DNA methylation markers for early detection of colorectal cancer

DNA methylation markers for early detection of colorectal cancer:

Clinical applicability and biological function

Layout: Tiny Wouters
Cover design: Eefje Meijer
Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress
Published by: Uitgeverij BOXPress. Vianen

ISBN: 9789462955295

Printing of this thesis was financial supported by Stichting PAMM and the University of Maastricht.

DNA methylation markers for early detection of colorectal cancer:

Clinical applicability and biological function

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, Prof. dr. Rianne M. Letschert,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen op
vrijdag 2 december 2016 om 10.00 uur

door

Marjolein Hendrina Francisca Maria Lentjes

Promotores

Prof. dr. M van Engeland
Prof. dr. A.P. de Bruïne

Copromotor

Dr. V. Melotte

Beoordelingscommissie

Prof. dr. A.A.M. Masclee (voorzitter)
Dr. A.J.C. van den Brule (Jeroen Bosch ziekenhuis te ‘s-Hertogenbosch)
Prof. dr. B. Kremer
Dr. B. Küsters (Radboud Universitair Medisch Centrum te Nijmegen)
Prof. dr. H.J.T. Rutten (Catharina ziekenhuis te Eindhoven / Maastricht Universitair Medisch Centrum)

Contents

Chapter 1

General introduction

Epidemiology

Colorectal cancer (CRC) is a major health problem worldwide. CRC is, after lung and prostate cancer, the third most diagnosed cancer in males and, after breast cancer, the second most diagnosed cancer in females. CRC is also one of the leading causes of cancer-related mortality.[1] An estimated 1,4 million new CRC cases per year and 693,900 CRC-related deaths occurred in 2012.[1] The highest incidence rates are in Australia/New Zealand, Europe and Northern America. In the Netherlands, each year more than 15,000 patients are diagnosed and over 5,000 patients die of CRC.[2] Incidence rates are lower in less developed countries and in certain parts of the world, amongst which Western Asia and Eastern Europe.[1] These geographic variations can be explained by the variable prevalence of risk factors for CRC, including alcohol consumption[3], smoking,[4] and dietary habits[5,6] such as high intake of processed or red meat, low fiber and whole-grain intake. In contrast to the increasing incidence, a decrease in mortality has been observed owing to a combination of improved therapies, reduced prevalence of risk factors and CRC screening.[7,8] However, there are still countries with increasing incidence and mortality rates, amongst which Brazil and Russia.[9,10]

Diagnosis, staging and therapy

The main routes by which individuals are diagnosed with CRC consist of population-based screening or clinical evaluation of symptoms, such as rectal bleeding, unexplained weight loss or fatigue. When CRC is suspected, both routes eventually lead to colonoscopy in which the colon is examined and discovered lesions can be biopsied or removed. The obtained tissue samples are microscopically analyzed by a pathologist and, if present, diagnosed with precancerous lesions, so-called adenomas which can be divided into low grade or high grade dysplastic lesions and into different subtypes (tubular, tubulovillous and villous type), or cancer.

Staging of CRC is important in order to estimate the patient prognosis and helps determining the most accurate therapy. As for most solid tumors, the TNM classification of the American Joint Committee on Cancer (AJCC) and the International Union Against Cancer (UICC) is the most commonly used system to classify CRC.[11] This system divides CRC into Stage I to Stage IV according to the extent of tumor spread. Determination of the stage is accomplished by combining three different components:
T – the depth of invasion of the primary Tumor in the intestinal wall
N – the absence or presence and extent of regional lymph Node metastasis
M – the absence or presence of distant Metastasis

In stage I disease generally only surgical excision of the tumor will be sufficient, whereas in stage III CRC adjuvant chemotherapy is indicated with addition of neoadjuvant radiotherapy in rectal cancer. Stage II CRCs are divided in low and high risk

subgroups. High risk stage II disease is identified through evaluation of several histopathological parameters, namely T4 invasion depth, less than 10 lymph nodes analyzed, obstruction or perforation of the colonic wall, presence of angio-invasion and poor differentiation of the tumor. In these stage II high risk CRCs, adjuvant chemotherapy should be considered.[12] Stage IV disease will be treated by extensive surgical resection with adjuvant radio- and/or chemotherapy or, in the case of non-treatable disease, cytotoxic chemotherapy with or without targeted therapy in a palliative setting.[13] At this moment, the most successful targeted therapy in the treatment of CRC is the administration of monoclonal antibodies against key signaling molecules such as bevacizumab, an anti-VEGF antibody[14], and cetuximab and panitumumab which target EGFR in *KRAS* wildtype CRCs.[15-17] Currently, immunotherapy, such as the inhibition of the checkpoints PD1 and PDL1, is under investigation for CRC, and shows good clinical response in a subset of CRC patients; the mismatch repair deficient CRCs.[18,19]

Survival rates are inversely correlated to the stage at CRC diagnosis, with 5-years survival rates decreasing from over 90% in stage I CRC to approximately 10% in stage IV disease.[20] This emphasizes the importance of early detection of colorectal cancer.

Colorectal cancer biology

CRC can be divided in hereditary, familial, and sporadic CRC. Several hereditary syndromes, such as Familial Adenomatous Polyposis (FAP), Lynch syndrome (or also known as Hereditary Non-Polyposis Colorectal Cancer; HNPCC), MUTYH- associated polyposis (MAP) and several hamartomatous conditions have been described in which individuals have an increased risk of CRC development. HNPCC and FAP are the most common form of hereditary CRCs. FAP is characterized by development of numerous colorectal adenomas and carries a 100% risk of developing CRC at relatively young age. FAP is caused by inherited germline mutations of the *adenomatous polyposis coli*, or *APC* gene.[21,22] The second copy of the *APC* allele is inactivated through mutation or less frequently by loss of heterozygosity (LOH).[23-25] Lynch syndrome is caused by an inherited germline mutation in one of the mismatch repair (MMR) genes, most frequently affecting the *MSH2* or *MLH1* gene.[21,22] The second allele of the mutated MMR gene in Lynch syndrome is lost through mutation, LOH or promoter hypermethylation of the *MLH1* gene.[23] Biallelic inactivation of MMR genes causes defective repair of single base mismatches and insertion/deletions during DNA replication, predominantly in regions with repetitive nucleotide sequences (known as microsatellites), and leads to microsatellite instability (MSI).[23] Lynch syndrome carries an approximately 80% risk of developing CRC.

The familial subtype constitutes 20-40% of CRC cases and is characterized by a two to three times higher chance of developing CRC in individuals of which first degree

relatives have suffered from sporadic CRC. So far, the genes associated with familial CRC have not been identified.

The vast majority of CRCs, between 60 and 80%, occurs sporadically and mainly develops through a multistep process beginning with the transformation of normal colon mucosal cells to an adenomatous intermediate and ultimately malignant transformation to cancer, also known as the adenoma-carcinoma sequence.[26,27] Adenomatous polyps can occur as polypoid or nonpolypoid lesions, the latter have been described to have a more aggressive behavior, underlining the importance of recognizing these lesions during colonoscopy.[28]

The adenoma-carcinoma sequence, described by Fearon and Vogelstein in 1990, was the first genetic model for CRC tumorigenesis and describes progressive accumulation of genetic alterations during CRC development.[23,26,27,29] Genetic alterations are found in three categories of genes: 1) tumor suppressor genes, such as *APC*, *DCC (Deleted* in *Colon Cancer)*, *TP53*, *SMAD2*, and *SMAD4*; 2) oncogenes, such as *KRAS* and *β-catenin*; and 3) DNA repair genes, such as MMR genes.[29,30] Alterations during neoplastic progression include activating mutations in *KRAS* and 18q LOH involving *SMAD2* and *SMAD4*.[29] The tumor suppressor *TP53* is mutated in 40 to 50% of colon carcinomas, but is uncommonly affected in adenomas, suggesting that this mutation occurs in a late stage of tumorigenesis.[26,31] There seems to be a preferred sequence of changes, although the accumulation of alterations is most critical, rather than their occurrence in a specific order.

In addition to genetic alterations, epigenetic changes have been described in CRC. The best characterized epigenetic alterations are: 1) DNA methylation; 2) histone modifications and 3) aberrant expression of noncoding RNAs (ncRNAs).[32] Epigenetic regulation refers to heritable and reversible changes in gene expression maintained through cell division, without changing the actual DNA sequence.[33,34] One of the most studied and best characterized epigenetic alterations is DNA methylation, which occurs at the fifth carbon of the pyrimidine ring of cytosines located in CpG dinucleotides. Promoter CpG island methylation is often associated with inhibition of gene expression.[35,36] This inhibition is a result of direct blocking of the target binding site of transcription factors, or indirectly by recruiting methylbinding domain proteins (MBDs) and histone modifying enzymes resulting in transcriptionally silent heterochromatin.[35,37] In CRC, numerous genes are shown to exhibit promoter hypermethylation in the same pathways affected by genetic alterations.[38-40] For example, according to Knudson's hypothesis the first hit is the loss of the *APC* gene early in the neoplastic process by a somatic mutation.[41] Next, the second hit is accomplished by inactivation of the intact copy of *APC* due to an inactivating somatic mutation or by promoter hypermethylation.[42,43]

Based on the type and extension of genomic alterations in CRC, three major pathways have been identified; chromosomal instability (CIN), microsatellite instability (MSI) and CpG island methylator phenotype (CIMP). CIN is characterized by non-

random gains or losses of whole or parts of chromosomes and is the most common type of genomic instability, comprising 65-70% of the sporadic CRCs. Known tumor suppressor genes in CRC, such as *APC* and *SMAD4* are often located in frequently deleted chromosomal regions and chromosomal gains are associated with activation of proto-oncogenes, such as *MYC* and *EGFR*.[44-46] In addition to the CIN pathway, about 10-15% of sporadic CRCs are characterized by MSI.[47] MSI is the condition of genetic hypermutability that results from a defective mismatch repair (MMR) system. In contrast to HNPCC, the MMR system in sporadic MSI CRCs is predominantly malfunctioning because of somatic epigenetic silencing of *MLH1*.[48] This results in accumulation of mutations in microsatellites and affects key signaling pathways when these mutated microsatellites are located in DNA coding regions of driver genes with functions such as DNA repair, signal transduction, cell cycle regulation, and apoptosis.[49,50] Almost all CRCs have either CIN or MSI, suggesting an important role for genomic instability in the carcinogenesis of CRC.[51] A subset of CRCs, approximately 20%, displays significantly more methylation than other CRCs and are designated CIMP. The CIMP phenotype is poorly defined and shows overlap with MSI tumors and partly with CIN tumors.[52,53] Weisenberger et al. divided CIMP into two categories, CIMP-positive and CIMP-negative by defining CIMP as CpG island promoter methylation of ≥3 out of the five Weisenberger markers (*CACNA1G*, *IGF2*, *NEUROG1*, *RUNX3* and *SOCS1*).[54] However, several other groups have suggested a division into three different groups, CIMP-high, CIMP-low and CIMP-negative.[55-58] Despite the lack of a consensus definition of CIMP[52] and the use of multiple gene panels, marker thresholds, and laboratory methods, it is clear that CIMP is associated with *BRAF* mutations and a worse survival in CRC.[39,59,60]

Using large scale sequencing approaches, Sjöblom et al. and Wood et al. identified a variety of mutated genes in CRC which together revealed a personal cancer genome rather than a simple common carcinogenesis pathway. The most frequently mutated genes in CRC identified in these studies were *APC*, *KRAS* and *TP53* next to many less frequently mutated genes, such as *MAP2*, *PHIP* and *ERCC6*.[61,62] More recently, the comprehensive multidimensional analysis performed by The Cancer Genome Atlas (TCGA) consortium confirmed the subclassification of hypermutated and non-hypermutated CRCs. The majority of hypermutated CRCs shows high levels of MSI, CIMP and/or *MLH1* promoter methylation. The non-hypermutated CRCs exhibit not only mutations in the known tumor-related genes, such as *APC* and *KRAS* but also in several other genes, such as *FAM123B* and *SOX9*. This difference between hypermutated and non-hypermutated CRCs may point to divergent sequences of genetic events during carcinogenesis.[44] Based on gene-expression analyses, Guinney et al. recently postulated a classification of sporadic CRCs into four main subtypes, the consensus molecular subtypes (CMSs). CMS1 is characterized by a hypermutational status, MSI, CIMP-high and upregulated immune response pathways whereas CMS2 predominantly shows somatic copy number changes and upregulation of WNT and MYC

signaling. CMS3 exhibits low somatic copy number alterations, frequent *KRAS* mutations, CIMP-low status and enrichment for multiple metabolism signatures. Finally, CMS4 displays upregulation of genes involved in the epithelial-mesenchymal transition and higher chromosomal instability.[63] This subdivision into molecular homogeneous subsets of sporadic CRCs might lead to a better understanding of CRC carcinogenesis, and thus a more adequate prediction of prognosis or therapeutic response.

Colorectal cancer screening

CRC development takes years and occurs through a multistep process via the adenoma-carcinoma sequence which provides a window of opportunity to detect CRC in an early stage. Because adenomas and early stage CRCs often present without clinical symptoms, an active search for colonic neoplasia is needed. CRC screening leads not only to the discovery of CRC in an earlier stage with reduction of morbidity and mortality but also to a lower CRC incidence.[8] Additionally, CRC screening is cost-effective or even cost-saving.[64] CRC screening occurs in the United States, Canada and several European countries, including the Netherlands. Two approaches for screening can be distinguished: organized screening and opportunistic screening. In organized screening, a selected group of individuals are invited to participate in screening and offered follow up if a positive screen is identified. In Europe, the majority of countries uses this approach. The opportunistic variant mainly occurs in the USA and consists of individuals whom actively ask for a screening test, or the test has been offered by a health care professional. Various methods of CRC screening are available such as colonoscopy, flexible sigmoidoscopy, and fecal occult blood tests (FOBTs). Screening using flexible sigmoidoscopy, colonoscopy and guaiac-based (g)FOBT is associated with a reduction in CRC mortality.[65-67] Colonoscopy has a high sensitivity and the discovered precancerous or cancerous lesions instantly can be removed or biopsied for pathological examination. However, it is an invasive method with a small risk of complications such as bleeding and bowel perforation, and well trained personnel is required making it an expensive screening method. The same accounts for flexible sigmoidoscopy in which only the sigmoid and rectum are examined. Different (non-invasive) FOBT techniques have been developed including the gFOBT and the fecal immunochemical test (FIT), detecting minimal amounts of blood in stool samples with sensitivities respectively ranging from 9-24% and 32-53% for detecting advanced neoplasia and 13-50% and 79% for detecting CRC.[68] At this moment, data from randomized trials are not available for the FIT, but because of the higher sensitivity and specificity of the FIT compared to the gFOBT, one can assume that this test will have a positive impact on CRC mortality.

More recent non-invasive tests include biomarker tests for stool and blood samples. Stool tests are based on the detection of cancer specific molecular alterations in CRC cells which exfoliate from the neoplastic surface. In blood tests, circulating genomic tumor material is used to detect CRC. The first study describing detection of gene mutations in tumor-derived DNA in stool was published by Sidransky et al. in 1992.[69] DNA mutations, epigenetic alterations and differential expressed ncRNAs are amongst the most promising biomarkers for the detection of CRC within stool and blood samples. Analyzing genetic aberrations[70-72], such as *KRAS*, *APC, BRAF, TP53* and/or MSI, led to the development of different stool DNA tests, including the commercial PreGen-Plus test of Exact Sciences[73-76] and a digital melt curve (DMC) assay[77]. Stool DNA tests have been shown to have higher CRC detection rates in comparison with the gFOBT.[78,79] The most recently developed, FDA approved, stool test is the Cologuard® (Exact Sciences) which consists of *KRAS* mutation analysis and detection of *NDRG4* and *BMP3* promoter CpG island methylation in combination with an immunochemical assay for human globin. This test detects up to 42% advanced adenomas and 92% CRCs in asymptomatic average-risk individuals.[80]

Aim and outline of the thesis

Currently, colonoscopy and gFOBT are the only screening methods which have been proven to lead to CRC mortality reduction.[65-67] However, the former is invasive and the latter can be improved regarding sensitivity and specificity. Though colonoscopy remains the gold standard, more sensitive and specific pre-screening tests should be developed in order to reduce the number of individuals who have to undergo this invasive and expensive detection method. More patient friendly tests, such as stool-based or blood-based DNA tests, are promising to improve the detection rates of CRC and high risk precursor lesions in comparison to the detection of globin in feces which depends on bleeding of the neoplastic tissue.

The **aim** of this thesis is to identify alterations in the CRC epigenome that can be used as biomarkers for the early detection of CRC. In addition, we explored the biological function of the identified genes in order to fully understand their potential role as a biomarker for early CRC detection.

In **chapter 2**, we describe the frequency of *GATA4* and *GATA5* promoter CpG island hypermethylation in CRC and its associations with clinicopathological characteristics. Additionally we show the potential biomarker use of *GATA4* promoter methylation in tumor-derived DNA in stool and *in vitro* evidence for a tumor suppressor role of these proteins in CRC. In **chapter 3** the current knowledge of GATA transcription factor functions in development and disease and the clinical applications of GATA factor

aberrations is reviewed and discussed. In **chapter 4** we identify *N-Myc downstream regulated gene 4* (*NDRG4*) as a potential biomarker for CRC detection in stool DNA and its potential tumor suppressor function. To further explore and understand the role of this gene in the gut, we evaluated the expression pattern of NDRG4 in the gut and in other tissues in **chapter 5**. Despite the proof of principle for the non-invasive detection of CRC-derived DNA in stool, we hypothesized that a blood-based assay, not depending on stool sampling, has potential for higher patient compliance. Therefore, in **chapter 6** we examined the performance of *NDRG4* and *GATA5* together with two novel genes, *spectrin repeat containing nuclear envelop 1* (*SYNE1*) and *forkhead box protein E1* (*FOXE1*) as methylation markers in plasma DNA for CRC detection. In **chapter 7** we summarize and discuss the data presented in this thesis and reflect on future studies for identification, implementation and biological function of DNA methylation markers for early detection of CRC.

References

1. Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. *CA Cancer J Clin* 2015;65(2):87-108.
2. Integraal Kankercentrum Nederland. 2015; Available from: www.iknl.nl.
3. Fedirko V, Tramacere I, Bagnardi V, et al. Alcohol drinking and colorectal cancer risk: an overall and dose-response meta-analysis of published studies. *Ann Oncol* 2011;22(9):1958-1972.
4. Botteri E, Iodice S, Bagnardi V, et al. Smoking and colorectal cancer: a meta-analysis. *JAMA* 2008;300(23):2765-2778.
5. Song M, Garrett WS, Chan AT. Nutrients, foods, and colorectal cancer prevention. *Gastroenterology* 2015;148(6):1244-1260 e16.
6. Bouvard V, Loomis D, Guyton KZ, et al. Carcinogenicity of consumption of red and processed meat. *Lancet Oncol* 2015;16(16):1599-1600.
7. American Cancer Society. Available from: www.cancer.org.
8. Edwards BK, Ward E, Kohler BA, et al. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer* 2010; 116(3):544-573.
9. Bosetti C, Bertuccio P, Malvezzi M, et al. Cancer mortality in Europe, 2005-2009, and an overview of trends since 1980. *Ann Oncol* 2013;24(10):2657-2671.
10. Chatenoud L, Bertuccio P, Bosetti C, et al. Trends in mortality from major cancers in the Americas: 1980-2010. *Ann Oncol* 2014;25(9):1843-1853.
11. Sonbin LH, Gospodarowicz M, Wittekind C. International Union Against Cancer TNM Classification of Malignant Tumours. 7 ed. 2009, Hoboken, NJ: Wiley-Blackwell.
12. Benson AB, 3rd, Schrag D, Somerfield MR, et al. American Society of Clinical Oncology recommendations on adjuvant chemotherapy for stage II colon cancer. *J Clin Oncol* 2004;22(16):3408-3419.
13. Richtlijnen oncologische zorg. Available from: www.oncoline.nl.
14. Saltz LB, Clarke S, Diaz-Rubio E, et al. Bevacizumab in combination with oxaliplatin-based chemotherapy as first-line therapy in metastatic colorectal cancer: a randomized phase III study. *J Clin Oncol* 2008;26(12):2013-2019.
15. Van Cutsem E, Kohne CH, Hitre E, et al. Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N Engl J Med* 2009;360:1408-1417.
16. Douillard JY, Oliner KS, Siena, S, et al. Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer. *N Engl J Med* 2013l369(11):1023-1034.
17. Roberts PJ, Stinchcombe TE. KRAS mutation: should we test for it, and does it matter? *J Clin Oncol* 2013;31(8):1112-1121.
18. Amin M, Lockhart AC. The potential role of immunotherapy to treat colorectal cancer. *Expert Opin Investig Drugs* 2015;24(3):329-344.
19. Le DT, Uram JN, Wang H, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* 2015;372(26):2509-2520.
20. Stedman MR, Feuer EJ, Mariotto AB. Current estimates of the cure fraction: a feasibility study of statistical cure for breast and colorectal cancer. Journal of the National Cancer Institute. *Monographs* 2014;2014(49):244-254.
21. Jasperson KW, Tuohy TM, Neklason DW, et al. Hereditary and familial colon cancer. *Gastroenterology* 2010;138(6):2044-2058.
22. Cheah PY. Recent advances in colorectal cancer genetics and diagnostics. *Crit Rev Oncol Hematol* 2009;69(1):45-55.
23. Fearon ER. Molecular genetics of colorectal cancer. *Annu Rev Pathol* 2011;6: 479-507.
24. Lamlum H, Ilyas M, Rowan A, et al. The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of the germline mutation: a new facet to Knudson's 'two-hit' hypothesis. *Nat Med* 1999;5(9):1071-1075.
25. Crabtree M, Sieber OM, Lipton L, et al. Refining the relation between 'first hits' and 'second hits' at the APC locus: the 'loose fit' model and evidence for differences in somatic mutation spectra among patients. *Oncogene* 2003;22(27): 4257-4265.

26. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990; 61(5):759-767.
27. Vogelstein B, Fearon ER, Hamilton SR, et al. Genetic alterations during colorectal-tumor development. *N Engl J Med* 1988;319(9):525-532.
28. Rembacken BJ, Fujii T, Cairns A, et al. Flat and depressed colonic neoplasms: a prospective study of 1000 colonoscopies in the UK. *Lancet* 2000;355(9211): 1211-1214.
29. Pritchard CC, Grady WM. Colorectal cancer molecular biology moves into clinical practice. *Gut* 2011;60(1):116-129.
30. Saif MW, Chu E. Biology of colorectal cancer. *Cancer J* 2010;16(3):196-201.
31. Iacopetta B. TP53 mutation in colorectal cancer. *Hum Mutat* 2003;21(3):271-276.
32. Ragusa, M., Barbagallo, C., Statello, L., et al., Non-coding landscapes of colorectal cancer. *World J Gastroenterol* 2015;21(41):11709-11739.
33. Migheli F, Migliore L. Epigenetics of colorectal cancer. *Clin Genet* 2012;81(4): 312-318.
34. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. *Cell* 2012;150(1):12-27.
35. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. Nature reviews. *Genetics* 2002;3(6):415-428.
36. Jones PA. The DNA methylation paradox. *Trends Genet* 1999;15(1):34-37.
37. Lopez-Serra L, Esteller M. Proteins that bind methylated DNA and human cancer: reading the wrong words. *Br J Cancer* 2008;98(12):1881-1885.
38. Suzuki H, Watkins DN, Jair KW, et al. Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. *Nat Genet* 2004;36(4):417-422.
39. van Engeland M, Derks S, Smits KM, et al. Colorectal cancer epigenetics: complex simplicity. *J Clin Oncol* 2011;29(10):1382-1391.
40. Herman JG, Umar A, Polyak K, et al. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc Natl Acad Sci U S A*. 1998;95(12):6870-6875.
41. Knudson AG. Chasing the cancer demon. *Annu Rev Genet* 2000;34:1-19.
42. Levy DB, Smith KJ, Beazer-Barclay Y, et al. Inactivation of both APC alleles in human and mouse tumors. *Cancer Res* 1994;54(22):5953-5958.
43. Esteller M, Sparks A, Toyota M, et al. Analysis of adenomatous polyposis coli promoter hypermethylation in human cancer. *Cancer Res* 2000;60(16):4366-4371.
44. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012;487(7407):330-337.
45. Xie T, D' Ario G, Lamb JR, et al. A comprehensive characterization of genome-wide copy number aberrations in colorectal cancer reveals novel oncogenes and patterns of alterations. *PloS One* 2012;7(7):e42001.
46. Rajagopalan H, Nowak MA, Vogelstein B, et al. The significance of unstable chromosomes in colorectal cancer. *Nat Rev Cancer* 2003;3(9):695-701.
47. Boland CR, Goel A. Microsatellite instability in colorectal cancer. *Gastroenterology* 2010;138(6):2073-2087 e3.
48. Cunningham JM, Christensen ER, Tester DJ, et al. Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. *Cancer Res* 1998;58(15): 3455-3460.
49. Takayama T, Miyanishi K, Hayashi T, et al. Colorectal cancer: genetics of development and metastasis. *J Gastroenterol* 2006;41(3):185-192.
50. Alhopuro P, Sammalkorpi H, Niittymaki I, et al. Candidate driver genes in microsatellite-unstable colorectal cancer. *Int J Cancer* 2012;130(7):1558-1566.
51. Derks S, Postma C, Carvalho B, et al. Integrated analysis of chromosomal, microsatellite and epigenetic instability in colorectal cancer identifies specific associations between promoter methylation of pivotal tumour suppressor and DNA repair genes and specific chromosomal alterations. *Carcinogenesis* 2008; 29(2):434-439.
52. Hughes LA, Khalid-de Bakker CA, Smits KM, et al. The CpG island methylator phenotype in colorectal cancer: progress and problems. *Biochim Biophys Acta* 2012;1825(1):77-85.
53. Hughes LA, Melotte V, de Schrijver J, et al. The CpG island methylator phenotype: what's in a name? C*ancer Res* 2013;73(19):5858-5868.

54. Weisenberger DJ, Siegmund KD, Campan M, et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* 2006;38(7):787-793.
55. Shen L, Toyota M, Kondo Y, et al. Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proc Natl Acad Sci U S A.* 2007;104(47):18654-18659
56. Barault L, Charon-Barra C, Jooste V, et al. Hypermethylator phenotype in sporadic colon cancer: study on a population-based series of 582 cases. *Cancer Res* 2008; 68(20):8541-6.
57. Anacleto C, Leopoldino AM, Rossi B, et al. Colorectal cancer "methylator phenotype": fact or artifact? *Neoplasia* 2005;7(4):331-335.
58. Ogino S, Kawasaki T, Kirkner GJ, et al. CpG island methylator phenotype-low (CIMP-low) in colorectal cancer: possible associations with male sex and KRAS mutations. *J Mol Diagn* 2006;8(5):582-588.
59. Juo YY, Johnston FM, Zhang DY, et al. Prognostic value of CpG island methylator phenotype among colorectal cancer patients: a systematic review and meta-analysis. *Ann Oncol* 2014;25(12):2314-2327.
60. Simons CC, Hughes LA, Smits KM, et al. A novel classification of colorectal tumors based on microsatellite instability, the CpG island methylator phenotype and chromosomal instability: implications for prognosis. *Ann Oncol* 2013. 24(8): p. 2048-2056.
61. Sjoblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science* 2006;314(5797):268-274.
62. Wood LD, Parsons DW, Jones S, et al. The genomic landscapes of human breast and colorectal cancers. *Science* 2007;318(5853):1108-1113.
63. Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med* 2015;21(11):1350-1356.
64. Lansdorp-Vogelaar I, Knudsen AB, Brenner H. Cost-effectiveness of colorectal cancer screening. *Epidemiol Rev* 2011;33:88-100.
65. Holme O, Loberg M, Kalager M, et al. Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: a randomized clinical trial. *JAMA* 2014; 312(6):606-615.
66. Hewitson P, Glasziou P, Watson E, et al. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. Am J Gastroenterol 2008;103(6):1541-1549.
67. Zauber AG, Winawer SJ, O'Brien MJ, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *N Engl J Med* 2012;366(8):687-696.
68. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. *Gut* 2015;64(10):1637-1649.
69. Sidransky D, Tokino T, Hamilton SR, et al. Identification of ras oncogene mutations in the stool of patients with curable colorectal tumors. *Science* 1992;256(5053): 102-105.
70. Deuter R, Muller O. Detection of APC mutations in stool DNA of patients with colorectal cancer by HD-PCR. Hum Mutat 1998;11(1):84-89.
71. Traverso G, Shuber A, Levin B, et al. Detection of APC mutations in fecal DNA from patients with colorectal tumors. *N Engl J Med* 2002;346(5):311-320.
72. Eguchi S, Kohara N, Komuta K, et al. Mutations of the p53 gene in the stool of patients with resectable colorectal cancer. *Cancer* 1996;77(8 Suppl):1707-1710.
73. Calistri D, Rengucci C, Bocchini R, et al. Fecal multiple molecular tests to detect colorectal cancer in stool. *Clin Gastroenterol Hepatol* 2003;1(5):377-383.
74. Ahlquist DA, Skoletsky JE, Boynton KA, et al. Colorectal cancer screening by detection of altered human DNA in stool: feasibility of a multitarget assay panel. *Gastroenterology* 2000;119(5):1219-1227.
75. Brand RE, Ross ME, Shuber AP. Reproducibility of a multitarget stool-based DNA assay for colorectal cancer detection. Am J Gastroenterol 2004;99(7):1338-1341.
76. Syngal S, Stoffel E, Chung D, et al. Detection of stool DNA mutations before and after treatment of colorectal neoplasia. *Cancer* 2006;106(2):277-283.
77. Zou H, Taylor WR, Harrington JJ, et al. High detection rates of colorectal neoplasia by stool DNA testing with a novel digital melt curve assay. *Gastroenterology* 2009; 136(2):459-470.
78. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. *N Engl J Med* 2004; 351(26):2704-2714.
79. Ahlquist DA, Sargent DJ, Loprinzi CL, et al. Stool DNA and occult blood testing for screen detection of colorectal neoplasia. *Ann Intern Med* 2008;149(7):441-450, W81.

80. A stool DNA test (Cologuard) for colorectal cancer screening. *JAMA* 2014;312(23): 2566.

Chapter 2

GATA4 and GATA5 are potential tumor suppressors and biomarkers in colorectal cancer

Debby M.E.I. Hellebrekers*, Marjolein H.F.M. Lentjes*, Sandra M. van den Bosch, Veerle Melotte, Kim A.D. Wouters, Kathleen L.J. Daenen, Kim M. Smits, Yoshimitsu Akiyama, Yasuhito Yuasa, Silvia Sanduleanu, Carolina A.J. Khalid-de Bakker, Daisy Jonkers, Matty P. Weijenberg, Joost Louwagie, Wim van Criekinge, Beatriz Carvalho, Gerrit A. Meijer, Stephen B. Baylin, James G. Herman, Adriaan P. de Bruïne, Manon van Engeland

Clin Cancer Res. 2009;15:3990-3997

**both authors contributed equally to this work*

Abstract

Purpose

The transcription factors GATA4 and GATA5 are involved in gastrointestinal development and are inactivated by promoter hypermethylation in colorectal cancer (CRC). Here, we evaluated GATA4/5 promoter methylation as potential biomarkers for non-invasive CRC detection, and investigated the role of GATA4/5 in CRC.

Experimental design

Promoter methylation of GATA4/5 was analyzed in colorectal tissue and fecal DNA from CRC patients and healthy controls using methylation-specific PCR. The potential function of GATA4/5 as tumor suppressors was studied by inducing GATA4/5 overexpression in human CRC cell lines.

Results

GATA4/5 methylation was observed in 70% (63/90) and 79% (61/77) of colorectal carcinomas, respectively, and was independent of clinicopathological features. Methylation frequencies in normal colon tissues from non-cancerous controls were 6% (5/88, GATA4, $p<0.001$) and 13% (13/100, GATA5, $p<0.001$). GATA4/5 overexpression suppressed colony formation ($p<0.005$), proliferation ($p<0.001$), migration ($p<0.05$), invasion ($p<0.05$), and anchorage-independent growth ($p<0.0001$) of CRC cells. Examination of GATA4 methylation in fecal DNA from two independent series of CRC patients and controls yielded a sensitivity of 71% (95% confidence interval [CI]=55%-88%) and specificity of 84% (95% CI=74%-95%) for CRC detection in the training set, and a sensitivity of 51% (95% CI=37%-65%) and specificity of 93% (95% CI=84%-100%) in the validation set.

Conclusions

Methylation of GATA4/5 is a common and specific event in colorectal carcinomas, and GATA4/5 exhibit tumor suppressive effects in colorectal cancer cells *in vitro*. GATA4 methylation in fecal DNA may be of interest for CRC detection.

Statement of translational relevance

Detection of aberrantly methylated tumor suppressor genes in stool DNA of CRC patients provides an attractive strategy for non-invasive and early detection of CRC. The current manuscript demonstrates that GATA4/5 promoter methylation is a early, frequent and specific event in CRC, independent of clinicopathological features. We also show that GATA4 methylation is a sensitive and specific biomarker for CRC detection in stool DNA. In addition, we partly unravelled the function of GATA4 and GATA5 in CRC, by demonstrating that these proteins suppress colony formation, proliferation, migration, invasion and anchorage-independent growth of CRC cells, indicating a tumor suppressor role of GATA4 and GATA5 in CRC. To our knowledge, this is the first study that 1) reports on the tumor suppressive effects of GATA4 and GATA5 in CRC, and 2) demonstrates that GATA4 methylation is a promising biomarker for early CRC screening in stool DNA.

Introduction

Early detection of colorectal cancer (CRC) and high-risk precursor lesions will improve cure rates.[1] The gold standard for CRC detection is colonoscopy, however due to its invasive nature, many patients refrain from undergoing colonoscopy. Therefore, non-invasive screening modalities to select patients at risk for CRC for colonoscopy, are needed. Currently, testing for the presence of fecal occult blood (FOBT) is used.[2,3] Despite its low sensitivity, FOBT has been shown to reduce the incidence and risk of CRC death when used programmatically.[1,2,4,5] A promising non-invasive CRC screening modality is the detection of CRC-specific genetic alterations in stool-derived DNA,[6-9] however it needs improvement in terms of sensitivity and cost effectiveness.

Promoter CpG island hypermethylation analysis of serum and feces has the potential to be used as a non-invasive test for the early diagnosis of (colorectal) cancers.[10-15] However, since promoter methylation is also associated with aging[16] and chronic inflammation[17-19], proper selection of methylation markers is crucial for sensitive and specific detection of CRC.

Transcription factors GATA4 and GATA5 play an essential role in the development and differentiation of the gastrointestinal tract and are suggested to be involved in CRC development.[20-23] However, the (tumor suppressor) function of these genes is poorly understood.

Here we examined promoter hypermethylation of GATA4/5 in large, well-characterized series of CRCs and non-cancerous colorectal mucosa, and compared GATA4/5 methylation frequencies with those of other genes functionally involved and frequently methylated in CRC. In addition, we investigated the function of GATA4/5 in human CRC cells by transfecting these cells with a GATA4/5 expression vectors and measuring colony formation, proliferation, migration, invasion and anchorage-independent growth. Finally, we evaluated the use of GATA4 methylation in fecal DNA as potential biomarker for early CRC detection.

Methods

Study population

Formalin-fixed, paraffin-embedded colorectal mucosa tissue of colorectal cancer (CRC) patients (n=102) and patients without cancer (n=230) over 50 years of age was retrospectively collected from the archive of the Department of Pathology of the Maastricht University Medical Center (Supplementary Figure S2.1, Supplementary Table S2.1, and Supplementary Methods). An additional, independent set of 716 paraffin-embedded CRCs was derived from patients participating in the prospective Netherlands Cohort Study on Diet and Cancer (NLCS).[24,25] Tissue samples were handled and analyzed

in a blinded fashion during collection, storage, DNA isolation and PCR analysis. This study was approved by the Medical Ethical Committee of the Maastricht University Medical Center.

Methylation-specific PCR, BRAF mutation and microsatellite instability analysis

A 5 µm section of each tissue block was stained with haematoxylin and eosin and revised by a pathologist (AdB). Five 20 µm sections were deparaffinated prior to DNA-isolation using the Puregene® DNA Isolation Kit (Gentra Systems, Qiagen). Promoter CpG island methylation of *GATA4*, *GATA5, APC*, *p14*^{ARF}, O^{6}-*MGMT*, *HLTF*, *p16*^{INK4A} and *RASSF1A* was determined by sodium bisulfite treatment of genomic DNA followed by methylation-specific PCR (MSP) as described elsewhere.[25,26] For primer sequences and MSP conditions, see Supplementary Table S2.2. For analysis of *BRAF* mutation and microsatellite instability, see Supplementary Methods.

Cell culture and transfections

Human HCT116 and RKO CRC cell lines were cultured in Dulbecco's MEM (DMEM; Invitrogen, Breda, the Netherlands) containing 10% fetal bovine serum (FBS; HyClone, Etten-Leur, The Netherlands).

Full length GATA4/5 cDNAs subcloned into the pcDNA3 vector were named pcDNA3-GATA4 and pcDNA3-GATA5. RKO cells were transfected using Lipofectamine 2000 Reagent (Invitrogen) according to the manufacturer's protocol. After selection for 2-3 weeks with 1 mg/ml geneticin (G418; Invitrogen), individual clones were isolated. RKO clones constitutively expressing GATA4/5 protein, named pc-GATA4-1 and pc-GATA5-1, were maintained in medium containing G418 (1 mg/ml) and used for further experiments. Three RKO clones constitutively expressing empty vector (pc-con-1, pc-con-2 and pc-con-3) were used, and results of these 3 clonal lines were averaged and named pc-con (1-3). HCT116 cells were transfected with the Nucleofector Kit V (Amaxa Biosystems, Gaithersburg, MD) using the manufacturer's guidelines. Although up to 30 single colonies were picked after 2-3 weeks selection with 400 µg/ml G418, HCT116 clonal lines constitutively expressing GATA4/5 protein could not be maintained. Therefore, HCT116 cells were transfected with control construct (empty vector; pcDNA3), pcDNA3-GATA4 or pcDNA3-GATA5, selected for 10 days with G418 (400 µg/ml), and these heterogeneous cell populations, named pc-con, pc-GATA4 and pc-GATA5, respectively, were used for further experiments using medium without G418. For real-time RT-PCR-, Western Blot-, colony formation-, cell proliferation-, migration-, invasion-, and anchorage-independent growth assays, see Supplementary Methods.

Collection of fecal DNA

Colonoscopy negative control stool samples (n=75) were obtained from a population of healthy subjects over 50 years of age who were screened within the framework of a workplace-based community CRC screening study at the Maastricht University Medical Center. The Medical Ethical Committee of the Maastricht University Medical Center and the Dutch Health Council approved the study. Stool samples from colonoscopy confirmed CRC patients (n=75) were collected at the VU University Medical Center in Amsterdam. For subjects characteristics, see Supplementary Table S2.3. Written informed consent was obtained for all stool samples. Control stool samples and a subset of CRC samples were collected within 2 weeks prior to colonoscopy. Some CRC stool samples were collected 5 to 7 days following colonoscopy. CRC stool samples were only collected when the tumor was not resected after colonoscopy. Stool samples were stored and processed (see Supplementary Methods section) in one center, and handled and analyzed in a blinded fashion during collection, storage, DNA isolation and PCR analysis.

Quantitative MSP

Quantitative MSP (qMSP) was applied on a 7900HT fast real-time PCR system (Applied Biosystems, Foster City, CA). The PCR reaction was carried out in 12 µl volume containing buffer (16.6 mM $(NH_4)2SO_4$, 67 mM Tris, 6.7 mM $MgCl_2$, 10 mM ß-mercaptoethanol), 5 mM dNTP, 6 ng/µl forward primer, 18 ng/µl reverse primer, 0.16 µM molecular beacon, 0.1 µg BSA, 0.4 units Jumpstart DNA Taq polymerase (Sigma Aldrich, Zwijndrecht, the Netherlands), and 2.4 µl DNA. Cycling parameters were: 5 min 95°C, followed by 45 cycles of 30 sec 95°C, 30 sec 57°C and 30 sec 72°C, followed by 5 min 72°C. A standard curve ($2*10^6$ - 20 copies) was included to determine copy numbers of unknown samples by interpolation of their Ct values to the standard curve. For primer- and molecular beacon sequences, see Supplementary Table S2.2. Receiver operator characteristic (ROC) curve analysis was used to assess the best cutoff value (an optimal cutoff value was determined by the point on the ROC curve closest to 100% specificity and corresponding to the highest sensitivity), and to determine diagnostic performance, using the Area Under the Curve (AUC). Positivity for *GATA4* methylation was considered if a methylation value was higher than the cutoff.

Statistical analysis

We used the Pearson's χ^2 or Fisher's Exact test and the One-way ANOVA, Kruskal-Wallis or Mann-Whitney test where appropriate to compare categorical and continuous patient data respectively. Paired samples within the group of cases were analyzed using the Mc Nemar test and the paired t-test to compare categorical and continuous data, respectively. Since significant differences in age and location of the tissue were

observed between CRC patients and controls (Supplementary Table S2.1), logistic regression analyses were used to adjust for age and location. Where appropriate, the Bonferroni method was used to correct for multiple comparisons. To examine sensitivity and specificity of every possible marker combination, WEKA System's Bayes Network machine learning was applied[27]. *In vitro* cell line experiments are given as mean values ± SEM. Analysis of cell growth curves was performed by means of the two-way ANOVA test. Student's t-test was used for analyses of ^{3}H-thymidine incorporation and anchorage-independent growth. Colony formation assay, quantitative real-time RT-PCR, migration, and invasion assays analyses were done using the Mann-Whitney rank sum test. All p-values are two-sided and p-values ≤0.05 were considered statistically significant. Statistical analysis was performed in SPSS 12.0.1.

Results

High frequencies of GATA4 and GATA5 methylation in colorectal carcinomas and adenomas

GATA4 methylation was detected in 70% (63/90) of colorectal carcinomas, while methylation of this gene was observed in only 5 of 88 (6%) of normal colorectal tissues from non-cancerous controls were methylated (Table 2.1, $p<2*10^{-11}$). Methylation frequencies of *GATA5* were 79% (61/77) in colorectal carcinoma tissues and 13% (13/100) in non-cancerous controls (Table 2.1, $p<3*10^{-14}$). *GATA4/5* were 86% concomitantly methylated and 83% concomitantly unmethylated in colorectal carcinomas ($p<0.01*10^{-17}$, data not shown). Since promoter methylation has been described in inflammatory conditions of the gastrointestinal tract,[17-19] inflamed colorectal mucosa of non-cancerous patients was added to the control group (Table 2.1). This did not significantly increase *GATA4/5* methylation (7% and 12%, respectively; Table 2.1).

Comparing frequencies of *GATA4/5* methylation with those of other genes reported to be frequently methylated in CRC (*APC*, *p14^{ARF}*, *O^{6}-MGMT*, *HLTF*, *p16^{INK4A}* and *RASSF1A*[28-30]) (Table 2.1) showed that *GATA4/5* perform best in terms of specificity and sensitivity, respectively (Table 2.1). Furthermore, Bayesian network analysis showed that the sensitivity of *GATA4/5* methylation alone did not improve by adding any of the other markers. In addition, no other combination of methylation markers has a higher sensitivity as compared to *GATA4/5* alone (data not shown). No correlation was observed between *GATA4/5* methylation and the V600E *BRAF* mutation (found in 13% of all carcinomas (13/98)) or microsatellite instability (MSI) status (found in 15% of all carcinomas (15/101)) (data not shown).

Table 2.1 **Methylation frequencies in colorectal carcinomas compared to non-cancerous colorectal mucosa (normal or inflamed tissue).**

	CRC+	CRC-		CRC-	
	carcinoma	normal	p-value* (<)	normal plus inflamed	p-value* (<)
GATA4	63/90 (70%)	5/88 (6%)	$2*10^{-11}$	8/119 (7%)	$4*10^{-14}$
GATA5	61/77 (79%)	13/100 (13%)	$3*10^{-14}$	16/129 (12%)	$3*10^{-16}$
APC	47/100 (47%)	24/103 (23%)	$2*10^{-3}$	32/132 (24%)	$9*10^{-4}$
p14^{ARF}	37/86 (43%)	15/105 (14%)	$2*10^{-4}$	19/134 (14%)	$4*10^{-5}$
O^{6}-MGMT	50/96 (52%)	21/105 (20%)	$2*10^{-5}$	30/139 (22%)	$9*10^{-6}$
HLTF	50/96 (52%)	19/103 (18%)	$7*10^{-5}$	21/134 (16%)	$2*10^{-6}$
p16^{INK4A}	59/95 (62%)	14/68 (21%)	$3*10^{-7}$	25/97 (26%)	$5*10^{-7}$
RASSF1A	25/100 (25%)	14/101 (14%)	ns	16/131 (12%)	$2*10^{-2}$

Methylation frequencies are represented as the number of methylated samples/(divided by) the total number of samples analyzed (percentage). Logistic regression was used to adjust for age and location. CRC+: colorectal cancer patients, CRC-: non-cancerous individuals, ns: not significant. * Bonferroni-corrected p-value.

A second independent series of CRCs (NLCS)[24, 25] confirms the high frequency of *GATA4* (65% (369/572)) and *GATA5* (74% (440/592)) methylation (data not shown). Furthermore, *GATA4/5* promoter methylation was not significantly associated with TNM stage, tumor location, sex, age at diagnosis, histological type or grade of differentiation in both series (Table 2.2 and data not shown).

Table 2.2 ***GATA4/5*** **methylation frequencies in relation to clinicopathological features of colorectal carcinoma tissue.**

		GATA4	*GATA5*
TNM stage	I	11/15 (73%)	13/15 (87%)
	II	21/32 (66%)	21/29 (74%)
	III	25/34 (74%)	19/25 (76%)
	IV	6/9 (67%)	8/8 (100%)
Tumor Location	Proximal	32/41 (78%)	31/36 (86%)
	Distal	31/47 (66%)	29/39 (74%)
Sex	Male	29/42 (69%)	29/37 (78%)
	Female	34/48 (71%)	32/40 (80%)
Age at diagnosis*	≤ mean	28/38 (74%)	26/34 (76%)
	> mean	35/52 (67%)	35/43 (81%)
Histological type	Adenocarcinoma	50/75 (67%)	49/65 (75%)
	Mucinous Carcinoma	13/15 (87%)	12/12 (100%)
Differentiation	Poor	6/8 (75%)	6/7 (86%)
	Moderate	52/71 (73%)	47/59 (80%)
	Well	5/11 (46%)	8/11 (73%)

No significant correlations were found. *Individuals are divided into two groups; those with an age smaller than or equal to the mean age of the study population and those with an age higher than the mean age.

GATA4/5 methylation frequencies of adenomas which developed synchronously or metachronously to the tumor (n=75) and adenomas obtained from non-cancerous patients (n=72; 10 year follow-up) did not show significant differences between these

two groups. (Supplementary Table S2.4). No association of *GATA4/5* methylation with grade of dysplasia was observed, but more *GATA4* methylation in tubulovillous as compared with tubular adenomas was observed ($p<0.0001$, data not shown). Frequencies of *GATA4/5* methylation were not significantly different between normal colon mucosa obtained from CRC patients and non-cancerous controls (Supplementary Table S2.4). No association was found between *GATA4/5* promoter methylation in normal non-cancerous tissue and age at biopsy, sex or location of the normal tissue (data not shown).

Adenoma samples from CRC patients exhibited significantly higher *GATA4/5* methylation frequencies than normal colon mucosa from these patients ($p<0.002$ and $p<0.0005$, respectively) (Supplementary Table S2.5). Significantly higher *GATA4/5* methylation frequencies were observed in CRCs when compared to normal colon from CRC patients ($p<4*10^{-7}$ and $p<2*10^{-7}$, respectively). Although more methylation of *GATA4/5* was observed in carcinomas than in adenomas from CRC patients, this was not statistically significant (Supplementary Table S2.5).

Reduced colony formation and proliferation of CRC cells by GATA4 and GATA5

The high frequency of *GATA4/5* promoter methylation, as well as the frequent loss of the *GATA4* locus (8p23.1-p22), suggests that silencing these genes might confer a selection advantage. Therefore, expression constructs harboring full length *GATA4/5* cDNA were introduced into RKO in which GATA4 is present but *GATA5* is silenced,[22] showing reduced numbers of G418-resistant colonies (86% and 76% reduction, respectively) compared to transfection of empty vector (Figure 2.1A, $p<0.005$). Comparable results were found when using HCT116, in which both GATA4/5 are absent[22] (Figure 2.1B).

Single colonies of RKO and HCT116 transfectants were picked and expanded. For RKO, one clone with constitutively higher levels of GATA4 mRNA and protein than control transfectants was obtained (pc-GATA4-1), and one clone stably expressing GATA5 (pc-GATA5-1) (Figure 2.2A). Cell growth of three RKO control clones constitutively expressing empty vector (pc-con-1, -2, -3) was comparable and averaged (pc-con (1-3)), and proliferation of the GATA4/5 clones was significantly reduced compared with control clones (Figure 2.2B and C). In contrast to RKO, HCT116 single cell clones stably expressing GATA4/5 protein could not be maintained. Therefore, these cells were transiently transfected and selected with G418 for 10 days, and then seeded for functional assays. GATA4 and GATA5 transfectants expressed mRNA and protein of GATA4 and GATA5, respectively, and showed significantly decreased proliferation (32% and 45% inhibition after 5 days, respectively) as compared to control transfectants (pc-con) (Supplementary Figure S2.2). The percentage of cells with sub-diploid DNA content was measured using flow cytometry, but no differences in apoptosis or total cell death were observed (data not shown).

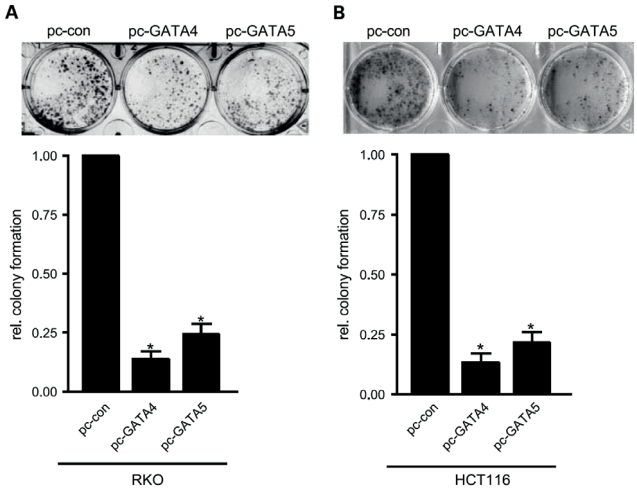

Figure 2.1 **GATA4 and GATA5 inhibit colony formation of human colorectal cancer cells.**
A, **B**. Colony formation of RKO (A) and HCT116 (B) cells transfected with a control vector (pc-con) or a GATA4 (pc-GATA4) or GATA5 (pc-GATA5) expression vector and selected for 2 weeks with G418. Quantification of colony formation is presented as mean values (± SEM) relative to control transfectants (pc-con) of three independent experiments (*p<0.005).

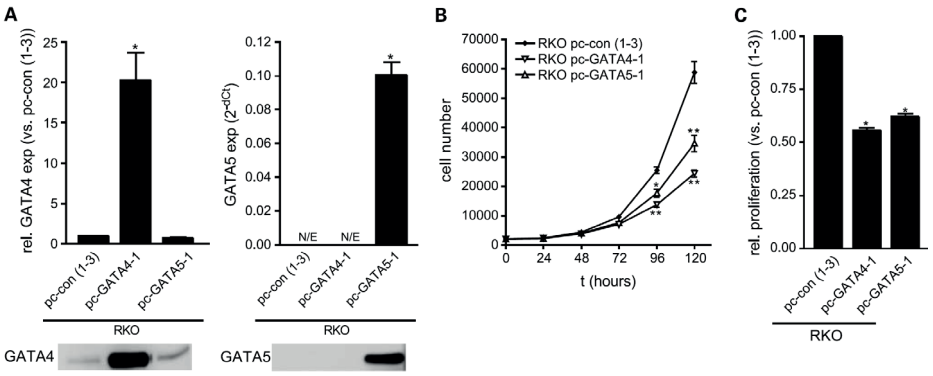

Figure 2.2 **GATA4 and GATA5 inhibit proliferation of human colorectal cancer cells.**
A. Bar graphs: GATA4 and -5 mRNA expression measured by real-time RT-PCR in RKO monoclonal cell lines. The pc-GATA4-1 and pc-GATA5-1 clones constitutively express GATA4 and -5, respectively, and pc-con (1-3) represents the average of 3 control (empty pcDNA3 vector) clones. Results are plotted as mean values (± SEM) of relative mRNA expression compared to pc-con (1-3) (GATA4) or as mean values (± SEM) of expression (calculated as 2^{-dCt}) (GATA5) of three independent experiments (*p<0.05 vs. pc-con (1-3)). N/E: not expressed. Gel images: Western Blot analysis of GATA4 and GATA5 protein in nuclear extracts of RKO cells. Only 1 out of 3 control clones are shown. **B**. Cell growth of RKO monoclonal cell lines. Results are plotted as mean values (± SEM) of cell numbers of three independent experiments (*p<0.01 vs. pc-con (1-3), **p<0.001 vs. pc-con (1-3)). **C**. Proliferation measured by 3H-thymidine incorporation. Data are expressed as mean relative proliferation values (± SEM) compared to pc-con (1-3) of three independent triplicate experiments (*p<0.0001 vs. pc-con (1-3)).

GATA4 and GATA5 suppress migration, invasion and anchorage-independent growth of CRC cells

We next examined the effects of GATA4/5 on migration and invasion of CRC cells using the modified Boyden chamber assay. Migration of GATA4/5 RKO clones was significantly lower when compared to control clones (Figure 2.3A, $p<0.05$). This was confirmed in HCT116, showing significantly decreased migration of the GATA4/5 transfected cells (68% and 73% inhibition, respectively) when compared to control transfectants (Figure 2.3B). Invasion of GATA4 and GATA5 RKO clones through matrigel-layered transwell membranes was also lower than that of control clones (Figure 2.3C, $p<0.05$). Similarly, the invasive activity of GATA4/5 HCT116 transfectants was also significantly reduced (87% and 74% inhibition, respectively) compared to empty vector transfected cells (Figure 2.3D and Supplementary Figure S2.3). Anchorage-independent growth of RKO monoclonal cell lines was assessed by soft agar colony formation. The number of colonies formed by GATA4/5 RKO clones was significantly lower (59% and 66% inhibition, respectively) compared to control clones (Supplementary Figure S2.4, $p<0.0001$). Furthermore, GATA/5 colonies were smaller than those produced by control transfectants (Supplementary Figure S2.4).

GATA4 methylation in fecal DNA as a potential biomarker for CRC detection

Since *GATA4* methylation was most specific (Table 2.1) and addition of *GATA5* did not significantly increase sensitivity compared to *GATA4* alone (data not shown), we further analyzed *GATA4* methylation in fecal DNA as a potential biomarker. Stool samples were collected from CRC patients (n=28), covering all stages of CRC, and 45 colonoscopy negative controls. *GATA4* methylation of fecal DNA was determined by quantitative MSP (qMSP). The Area Under the Curve (AUC) in the receiver operator characteristic (ROC) curve was 81% (95% confidence interval [CI] = 70%-89%) (Figure 2.4). The optimal *GATA4* methylation cutoff value was 8.1 (Figure 2.4). Using this cutoff, *GATA4* methylation was detected in fecal DNA from 20 of 28 patients and in 7 of the 45 control individuals, yielding a sensitivity of 71% (95% CI = 55%-88%) and a specificity of 84% (95% CI = 74%-95%). Since the mean age of the cases and controls differed significantly, ROC-GLM regression analysis was used to assess the accuracy of *GATA4* promoter methylation after adjustment for age.[31] Age did not significantly influence the accuracy ($p=0.71$, ROC-GLM regression model).

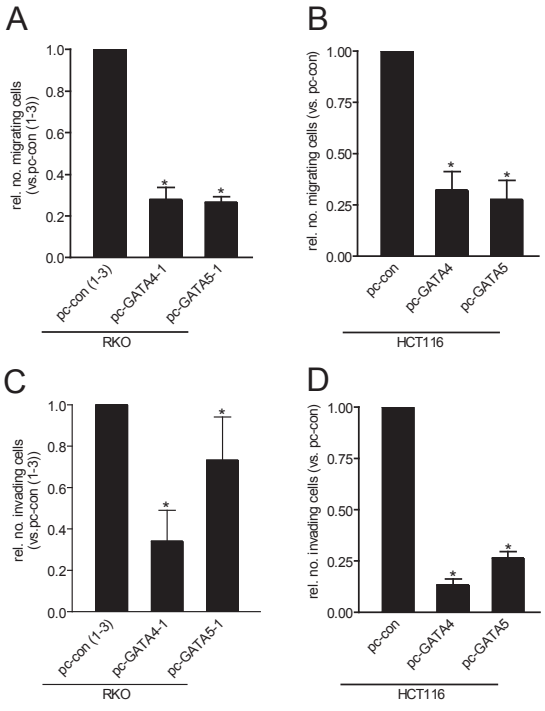

Figure 2.3 **GATA4 and GATA5 decrease migration and invasion of human colorectal cancer cells.**
A, **B**. Migration of RKO cell clones (A) and HCT116 cells (B) through transwells without matrigel, measured by direct counting of trespassed cells. Data are presented as mean relative numbers (± SEM) of migrated cells from several fields (200x) of two independent experiments (*p<0.05 vs. pc-con (1-3) (A) or vs. pc-con (B)). **C**, **D**. Invasion of RKO cell clones (C) and HCT116 cells (D) through transwells with matrigel. Results represent mean relative counts (± SEM) of trespassed cells from several fields (200x) of two independent experiments (*p<0.05 vs. pc-con (1-3) (C) or vs. pc-con (D)).

Sensitivity and specificity (using a *GATA4* cutoff value 8.1) was validated in an independent set of stool samples from CRC patients (n=47) and controls (n=30). This resulted in a sensitivity of 51% (95% CI = 37%-65%) and a specificity of 93% (95% CI = 84%-100%) of *GATA4* methylation in fecal DNA.

Sensitivity of *GATA4* promoter methylation in fecal DNA for detecting early stage (TNM stage I and II) and advanced stage (TNM stage III and IV) CRC was 10/18 (55%; early stage) versus 10/10 (100%; advanced stage) for the training set, and 14/29 (48%; early stage) versus 10/17 (59%; advanced stage) in the validation set. Although this pilot study shows proof of principle for detecting *GATA4* promoter methylation in stool, it appears that early stage CRCs shed less DNA when compared to advanced stage CRCs

which emphasizes the need for sensitive assays to isolate/capture DNA from early stage CRCs.

For a subset (n=19) of cases of which fecal DNA was examined for *GATA4* promoter CpG island methylation, the matching formalin-fixed paraffin embedded primary tumor tissue was available. *GATA4* promoter methylation was detected in 16 of 19 primary CRCs, and 10 of these 16 CRCs also exhibited methylation in the matched stool samples, yielding an analytical sensitivity of 63% (95% CI = 39%-86%; data not shown).

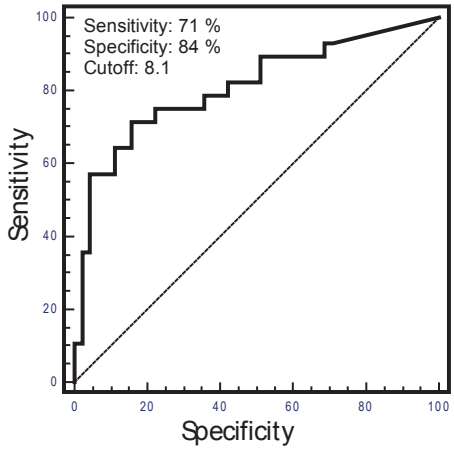

Figure 2.4 ***GATA4* methylation in fecal DNA as a biomarker for CRC detection.**
Receiver operator characteristic (ROC) curve for *GATA4* methylation considering 28 CRC stool samples and 45 normal control stool samples. The ROC curve displays the estimated sensitivity and specificity at various cutoff values for defining a positive test for *GATA4* qMSP. The determined optimal cutoff value for *GATA4* methylation was 8.1.

Discussion

Loss of *GATA4/5* expression due to promoter hypermethylation has been reported in primary colorectal, gastric, esophageal, lung, ovarian, and pancreatic (*GATA5* only) cancer.[22,32-35] To analyze the potential of *GATA4/5* as methylation markers for detection of CRC, we analyzed large series of CRC patients and controls and showed that methylation of *GATA4* (70%) and *GATA5* (79%) occurs at high frequencies in CRCs and at low levels in normal colorectal mucosa (6 and 13%, respectively). Methylation frequencies of *GATA4/5* are not increased in inflammatory colorectal tissues. *GATA4/5* methylation is highly prevalent in colorectal adenomas suggesting that methylation of *GATA4/5* is an early event in colorectal carcinogenesis. Lack of association of *GATA4/5*

methylation with clinicopathological characteristics indicates that *GATA4/5* methylation may be equivalently sensitive to early- and late stage CRC, and to proximal as well as distal CRC, thereby covering all CRC phenotypes, including microsatellite instable (MSI) and chromosomal instable tumors. These findings indicate that methylation of *GATA4/5* may be suitable markers for early diagnosis of CRC.

Methylation analysis of six other genes frequently and functionally methylated in CRC[28-30] showed frequent methylation of *APC*, *p16^{INK4A}*, and *O^6-MGMT* in normal and inflamed colorectal mucosa. Whether methylation of *APC* and *p16^{INK4A}* in normal colorectal mucosa represents a field effect (and thus a prognostic marker) as was published for *O^6-MGMT*[36] is not clear from this study.

Well-defined molecular markers will be helpful for non-invasive early diagnosis of CRC and might reduce mortality from this disease. The combined sensitivity and specificity of *GATA4* methylation for CRC detection compares well to other fecal DNA methylation markers such as *SFRP2*, vimentin and *HIC1*.[10,11,15] Nevertheless, increasing the sensitivity of *GATA4* methylation in fecal DNA is required to increase the applicability of this screening test. A higher sensitivity for the stool *GATA4* MSP test could be achieved using optimal isolation protocols for fecal DNA. For example, using methyl binding domain (MBD) protein columns to capture methylated DNA, which have been shown to markedly increase sensitivity without decreasing specificity[37], could be interesting in this respect. Also, identification of complementary (epi)genetic markers is required in order to obtain a multigene assay to augment diagnostic accuracy of fecal DNA testing. Machine learning revealed that neither addition of any of the other genes we tested nor a different gene panel outperforms the sensitivity of *GATA4* and *GATA5* methylation in primary colorectal carcinomas. This indicates that none of the other genes were complementary to the *GATA4/5* markers in primary colorectal carcinomas, and suggests the existence of a subset of CRCs with extensive promoter methylation and a subset without methylation of the markers tested in this study. When comparing *GATA4* methylation in stool DNA with the corresponding tumor tissue, two cases were identified in which *GATA4* methylation was found in stool DNA in the absence of methylation in the associated CRC tissue. This discrepancy might be due to the stool sample containing tumor cells from an area separate from where the tissue DNA was extracted, reflecting heterogeneity of *GATA4* promoter methylation in the tumor, or might be derived from additional tumors located upstream in the gastrointestinal tract such as esophageal or gastric tumors. In addition, the analytical sensitivity of 63% reveals that some of the patients with *GATA4* methylation in the primary CRC lacked methylation in the stool, which might result from the situation that detectable amounts of tumor cells may not have shed into the feces when it was collected.

For methylation analysis of stool DNA, qMSP was the method of choice, because this approach allows robust and sensitive automated analysis of clinical samples for use in molecular screening approaches, and specificity of this approach is enhanced by using labeled internal probes.

GATA4/5 have been implicated in cancer development, in which they would behave as tumor suppressors by activating the promoters of anti-tumor genes.[22,38] However, to our knowledge, tumor suppressive effects of these genes have never been reported in CRC, but only in GATA4-transfected ovarian tumor cells.[38] Here, we show that introduction of GATA4/5 into human CRC cell lines by transient and stable transfection results in inhibition of colony formation, cell growth, migration, invasion and anchorage-independent growth *in vitro*, suggesting that these genes are relevant tumor suppressor genes in CRC. However, the downstream target genes of GATA4/5 inducing the above mentioned effects remain to be identified.

In conclusion, we found that GATA4/5 exhibit tumor suppressive activities in CRC cells *in vitro* and show that promoter hypermethylation of *GATA4/5* is frequent and specific in primary CRCs. *GATA4* methylation in fecal DNA has potential to be used in a biomarker panel for improving pre-selection tests for colonoscopy.

References

1. Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM, Ederer F. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. *N Engl J Med* 1993;328:1365-1371.
2. Kronborg O, Fenger C, Olsen J, Jorgensen OD, Sondergaard O. Randomised study of screening for colorectal cancer with faecal-occult-blood test. *Lancet* 1996;348:1467-1471.
3. Winawer SJ, Zauber AG, Ho MN, O'Brien MJ, Gottlieb LS, Sternberg SS, Waye JD, Schapiro M, Bond JH, Panish JF, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N Engl J Med* 1993;329:1977-1981.
4. Hardcastle JD, Chamberlain JO, Robinson MH, Moss SM, Amar SS, Balfour TW, James PD, Mangham CM. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet* 1996;348:1472-1477.
5. Kewenter J, Brevinge H, Engaras B, Haglind E, Ahren C. Follow-up after screening for colorectal neoplasms with fecal occult blood testing in a controlled trial. *Dis Colon Rectum* 1994;37:115-119.
6. Traverso G, Shuber A, Levin B, Johnson C, Olsson L, Schoetz DJ, Jr., Hamilton SR, Boynton K, Kinzler KW, Vogelstein B. Detection of APC mutations in fecal DNA from patients with colorectal tumors. *N Engl J Med* 2002;346:311-320.
7. Dong SM, Traverso G, Johnson C, Geng L, Favis R, Boynton K, Hibi K, Goodman SN, D'Allessio M, Paty P, Hamilton SR, Sidransky D, Barany F, Levin B, Shuber A, Kinzler KW, Vogelstein B, Jen J. Detecting colorectal cancer in stool with the use of multiple genetic targets. *J Natl Cancer Inst* 2001;93:858-865.
8. Imperiale TF, Ransohoff DF, Itzkowitz SH, Turnbull BA, Ross ME. Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. *N Engl J Med* 2004;351:2704-2714.
9. Diehl F, Schmidt K, Durkee KH, Moore KJ, Goodman SN, Shuber AP, Kinzler KW, Vogelstein B. Analysis of mutations in DNA isolated from plasma and stool of colorectal cancer patients. *Gastroenterology* 2008;135:489-498.
10. Chen WD, Han ZJ, Skoletsky J, Olson J, Sah J, Myeroff L, Platzer P, Lu S, Dawson D, Willis J, Pretlow TP, Lutterbaugh J, Kasturi L, Willson JK, Rao JS, Shuber A, Markowitz SD. Detection in fecal DNA of colon cancer-specific methylation of the nonexpressed vimentin gene. *J Natl Cancer Inst* 2005;97:1124-1132.
11. Muller HM, Oberwalder M, Fiegl H, Morandell M, Goebel G, Zitt M, Muhlthaler M, Ofner D, Margreiter R, Widschwendter M. Methylation changes in faecal DNA: a marker for colorectal cancer screening? *Lancet* 2004;363:1283-1285.
12. Sanchez-Cespedes M, Esteller M, Wu L, Nawroz-Danish H, Yoo GH, Koch WM, Jen J, Herman JG, Sidransky D. Gene promoter hypermethylation in tumors and serum of head and neck cancer patients. *Cancer Res* 2000;60:892-895.
13. Belinsky SA, Liechty KC, Gentry FD, Wolf HJ, Rogers J, Vu K, Haney J, Kennedy TC, Hirsch FR, Miller Y, Franklin WA, Herman JG, Baylin SB, Bunn PA, Byers T. Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. *Cancer Res* 2006;66:3338-3344.
14. Hoque MO, Begum S, Topaloglu O, Chatterjee A, Rosenbaum E, Van Criekinge W, Westra WH, Schoenberg M, Zahurak M, Goodman SN, Sidransky D. Quantitation of promoter methylation of multiple genes in urine DNA and bladder cancer detection. *J Natl Cancer Inst* 2006;98:996-1004.
15. Lenhard K, Bommer GT, Asutay S, Schauer R, Brabletz T, Goke B, Lamerz R, Kolligs FT. Analysis of promoter methylation in stool: a novel method for the detection of colorectal cancer. *Clin Gastroenterol Hepatol* 2005;3:142-149.
16. Issa JP, Ottaviano YL, Celano P, Hamilton SR, Davidson NE, Baylin SB. Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nat Genet* 1994;7:536-540.
17. Sato F, Harpaz N, Shibata D, Xu Y, Yin J, Mori Y, Zou TT, Wang S, Desai K, Leytin A, Selaru FM, Abraham JM, Meltzer SJ. Hypermethylation of the p14(ARF) gene in ulcerative colitis-associated colorectal carcinogenesis. *Cancer Res* 2002;62:1148-1151.
18. Issa JP, Ahuja N, Toyota M, Bronner MP, Brentnall TA. Accelerated age-related CpG island methylation in ulcerative colitis. *Cancer Res* 2001;61:3573-3577.

19. Hsieh CJ, Klump B, Holzmann K, Borchard F, Gregor M, Porschen R. Hypermethylation of the p16INK4a promoter in colectomy specimens of patients with long-standing and extensive ulcerative colitis. *Cancer Res* 1998;58:3942-3945.
20. Molkentin JD. The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. *J Biol Chem* 2000;275:38949-38952.
21. Gao X, Sedgwick T, Shi YB, Evans T. Distinct functions are implicated for the GATA-4, -5, and -6 transcription factors in the regulation of intestine epithelial cell differentiation. *Mol Cell Biol* 1998;18:2901-2911.
22. Akiyama Y, Watkins N, Suzuki H, Jair KW, van Engeland M, Esteller M, Sakai H, Ren CY, Yuasa Y, Herman JG, Baylin SB. GATA-4 and GATA-5 transcription factor genes and potential downstream antitumor target genes are epigenetically silenced in colorectal and gastric cancer. *Mol Cell Biol* 2003;23:8429-8439.
23. Fujiwara Y, Emi M, Ohata H, Kato Y, Nakajima T, Mori T, Nakamura Y. Evidence for the presence of two tumor suppressor genes on chromosome 8p for colorectal carcinoma. *Cancer Res* 1993;53:1172-1174.
24. Brink M, de Goeij AF, Weijenberg MP, Roemen GM, Lentjes MH, Pachen MM, Smits KM, de Bruine AP, Goldbohm RA, van den Brandt PA. K-ras oncogene mutations in sporadic colorectal cancer in The Netherlands Cohort Study. *Carcinogenesis* 2003;24:703-710.
25. van Engeland M, Weijenberg MP, Roemen GM, Brink M, de Bruine AP, Goldbohm RA, van den Brandt PA, Baylin SB, de Goeij AF, Herman JG. Effects of dietary folate and alcohol intake on promoter methylation in sporadic colorectal cancer: the Netherlands cohort study on diet and cancer. *Cancer Res* 2003;63:3133-3137.
26. Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A.* 1996;93:9821-9826.
27. Witten IH, Frank E, Kaufmann M. *Data Mining: Practical Machine Learning Tools With Java Implementations*. San Francisco (CA): Morgan Kaufmann Publishers; 2000.
28. Esteller M, Corn PG, Baylin SB, Herman JG. A gene hypermethylation profile of human cancer. *Cancer Res* 2001;61:3225-3229.
29. van Engeland M, Roemen GM, Brink M, Pachen MM, Weijenberg MP, de Bruine AP, Arends JW, van den Brandt PA, de Goeij AF, Herman JG. K-ras mutations and RASSF1A promoter methylation in colorectal cancer. *Oncogene* 2002;21:3792-3795.
30. Moinova HR, Chen WD, Shen L, Smiraglia D, Olechnowicz J, Ravi L, Kasturi L, Myeroff L, Plass C, Parsons R, Minna J, Willson JK, Green SB, Issa JP, Markowitz SD. HLTF gene silencing in human colon cancer. *Proc Natl Acad Sci U S A.* 2002;99:4562-4567.
31. Janes H, Longton GM, Pepe M. Accommodating Covariates in ROC Analysis. *UW Biostatistics Working Paper Series* 2008;Working paper 322.
32. Guo M, Akiyama Y, House MG, Hooker CM, Heath E, Gabrielson E, Yang SC, Han Y, Baylin SB, Herman JG, Brock MV. Hypermethylation of the GATA genes in lung cancer. *Clin Cancer Res* 2004;10:7917-7924.
33. Guo M, House MG, Akiyama Y, Qi Y, Capagna D, Harmon J, Baylin SB, Brock MV, Herman JG. Hypermethylation of the GATA gene family in esophageal cancer. *Int J Cancer* 2006;119:2078-2083.
34. Wakana K, Akiyama Y, Aso T, Yuasa Y. Involvement of GATA-4/-5 transcription factors in ovarian carcinogenesis. *Cancer Lett* 2006;241:281-288.
35. Fu B, Guo M, Wang S, Campagna D, Luo M, Herman JG, Iacobuzio-Donahue CA. Evaluation of GATA-4 and GATA-5 methylation profiles in human pancreatic cancers indicate promoter methylation patterns distinct from other human tumor types. *Cancer Biol Ther* 2007;6(10).
36. Shen L, Kondo Y, Rosner GL, Xiao L, Hernandez NS, Vilaythong J, Houlihan PS, Krouse RS, Prasad AR, Einspahr JG, Buckmeier J, Alberts DS, Hamilton SR, Issa JP. MGMT promoter methylation and field defect in sporadic colorectal cancer. *J Natl Cancer Inst* 2005;97:1330-1338.
37. Zou H, Harrington J, Rego RL, Ahlquist DA. A novel method to capture methylated human DNA from stool: implications for colorectal cancer screening. *Clinl Chem* 2007;53:1646-1651.
38. Capo-chichi CD, Roland IH, Vanderveer L, Bao R, Yamagata T, Hirai H, Cohen C, Hamilton TC, Godwin AK, Xu XX. Anomalous expression of epithelial differentiation-determining GATA factors in ovarian tumorigenesis. *Cancer Res* 2003;63:4967-4977.

Supplementary methods

Study population

Collection and use of archival tissue for this study was approved by the Medical Ethical Committee (MEC) of the Maastricht University and the University Hospital Maastricht. Formalin-fixed, paraffin-embedded colorectal mucosa tissue of colorectal cancer (CRC) patients (n=102) and patients without cancer (n=230) over 50 years of age (diagnosed between 1995-2003) was retrospectively collected from the archive of the Department of Pathology of the University Hospital Maastricht (Supplementary Figure S2.1). Patient characteristics are shown in Supplementary Table S2.1. CRC was classified according to location as proximal colon (cecum through transverse colon) or distal colon (splenic flexure through sigmoid colon and rectum). If present, also normal colon tissue (n=94) and synchronous or metachronous adenomas (n=75) were collected from CRC patients. The non-cancerous patient group consists of (1) histologically normal biopsy material from patients undergoing endoscopy because of non-specific abdominal complaints (n=124, diagnosed between 1987-2004) and who did not develop adenomas or CRC, (2) adenoma biopsies (n=72, diagnosed between 1988-1995) from patients who did not develop CRC within 10 years, and (3) resected colon mucosa of patients with various inflammatory bowel conditions (n=34, diagnosed between 1985-2004) who did not develop adenomas or CRC. This last group includes Crohn's disease (n=1), colitis ulcerosa (n=6), diverticulitis (n=18), and non-specific inflammation (n=9). CRC patients are significantly older than non-cancerous patients ($p<0.001$) (Supplementary Table S2.1). The location of the tissue from non-cancerous patients obtained by biopsy (mainly distal colon) is significantly different from the location of the tissue obtained from CRC patients ($p<0.001$) (Supplementary Table S2.1). CRC and non-cancerous patients were excluded if being diagnosed with additional cancers other than non-melanoma skin cancer.

Detection of BRAF mutation

The common *BRAF* V600E mutation in exon 15 was analyzed by a semi-nested PCR and subsequent RFLP analysis as described previously[1]. Primer sequences and PCR conditions are available upon request.

Microsatellite instability analysis

The microsatellite instability (MSI) status of the CRCs was evaluated by amplification of *BAT-26*, *BAT-25*, *NR21*, *NR22* and *NR24* mononucleotide repeat fragments which were previously shown to be highly sensitive and specific for MSI[2]. Primer sequences were as described previously by Suraweera et al[2]. Amplified PCR products were analyzed on an

ABI Prism 3100 genetic analyzer. Allelic sizes were estimated using Genemapper version 4.0 software (Applied Biosystems).

Real-time RT-PCR

Total RNA was isolated using the GenElute Mammalian total RNA miniprep kit (Sigma Aldrich) according to the supplier's protocol. Possible genomic DNA contaminations were removed by on column DNAse treatment with the RNase-free DNAse set (Qiagen). Total RNA (1 μg) was reverse transcribed using the I-Script cDNA synthesis kit (Bio-Rad, Veenendaal, The Netherlands). Real-time RT-PCR was performed as described previously[3] using iQ SYBR Green Supermix (Bio-Rad). Primer sequences are listed in Supplementary Table S2.2.

Western Blot

Nuclear extracts were prepared using NE-PER kit (Pierce, Rockford, IL, USA) containing protease inhibitors (Complete protease inhibitor cocktail tablets; Roche Applied Science, Mannheim, Germany). Protein samples were separated by 10% SDS-PAGE and transferred to Protran nitrocellulose membrane (Schleicher and Schuell, Dassel, Germany). After blocking with 5% nonfat dry milk in 0.1% Tween 20 in PBS, the membranes were incubated with monoclonal mouse anti-GATA4 (1:200; Santa Cruz Biotechnology, Santa Cruz, CA) or monoclonal mouse anti-GATA5 (0.5 μg/ml; R&D systems, Wiesbaden, Germany) antibodies. After washing, the blots were incubated with horseradish peroxidase-conjugated secondary antibody and detected with the Enhanced Chemiluminescence System (Pierce).

Colony formation assay

CRC cells were transfected with pcDNA3, pcDNA3-GATA4 or pcDNA3-GATA5 vector as described above. The next day, cells were diluted 1:10 and G418 (RKO 1 mg/ml; HCT116 400 μg/ml) was added. After 14 days of selection, colonies were stained and quantified. Colony formation was assessed in three independent experiments.

Cell proliferation assay

Cells were seeded onto 96-well plates (2000 cells/well) and cell numbers were counted after 24, 48, 72, 96 and 120 hours. During the last 6 hours of the assay, the culture was pulsed with 0.3 μCi [methyl-^{3}H] thymidine (Amersham Life Science, Roosendaal, The Netherlands) per well. Activity was measured using liquid scintillation. Three independent experiments were performed and in each experiment, measurements were done in triplicate.

In vitro migration and invasion assays

Cell migration and invasion assays were performed using 24-well transwells (8 µm pore size) coated with (invasion) or without (migration) matrigel (BD Biosciences, Franklin Lakes, NJ). $25*10^4$ cells in 1% FBS-DMEM were seeded into the upper chamber, and DMEM containing 20% FBS was placed in the lower chamber. After 48 hours, cells on the lower surface of the membrane were fixed with methanol and stained with 1% Toluine Blue in 1% borax. Cells in several random microscopic fields (200 x magnification) from two independent experiments were counted.

Anchorage-independent growth assay

RKO cells were suspended in DMEM containing 0.35% agar, 10% FBS, 50 ng/ml streptomycin and 50 U/ml penicillin, and layered on DMEM containing 0.5% agar, 10% FBS, 50 ng/ml streptomycin and 50 U/ml penicillin in 6-well plates. Colonies were scored after two weeks of growth in three independent experiments in triplicate.

Preparation of fecal DNA

Stool samples were homogenized in stool homogenization buffer (Exact Sciences, Marlborough, MA, USA), aliquoted in portions containing the equivalent of 4g stool each and processed within 48 hours after defecation. After RNase A treatment, total DNA was precipitated and resuspended in 4 ml 1x TE. Half of the volume from this resuspended DNA was then used as input for the QIAamp DNA stool midi test kit (user developed protocol). 1.5 ml ASL buffer and an InhibitEX tablet were added to 2 ml sample. After mixing, incubation for one minute, and centrifugation, 150 µl proteinase K was added to 2 ml supernatant which was then mixed with 2.4 ml buffer AL and incubated for 10 minutes at 70°C. In order to bind the DNA to the column, the proteinase K treated sample was mixed with 2 ml ethanol (96-100%) and then loaded onto the column in portions of 3.3 ml maximum. The column was washed twice with wash buffers AW1 and AW2, after which the column was dried by centrifugation. The DNA was eluted by pipetting 200 µl buffer AE onto the column. After centrifugation, the eluate was loaded back onto the column and centrifuged again. Finally, 2 µg DNA was subjected to bisulfite modification in 96-wells format on a pipetting robot (Tecan), using the EZ-96DNA Methylation kit (Zymo Research, Orange, CA), according to the manufacturer's protocol.

1. Sieben NL, Roemen GM, Oosting J, et al: Clonal analysis favours a monoclonal origin for serous borderline tumours with peritoneal implants. J Pathol 210:405-11, 2006
2. Suraweera N, Duval A, Reperant M, et al: Evaluation of tumor microsatellite instability using five quasimonomorphic mononucleotide repeats and pentaplex PCR. Gastroenterology 123:1804-11, 2002
3. Thijssen VL, Brandwijk RJ, Dings RP, et al: Angiogenesis gene expression profiling in xenograft models to study cellular interactions. Exp Cell Res 299:286-93, 2004

Supplemental figures

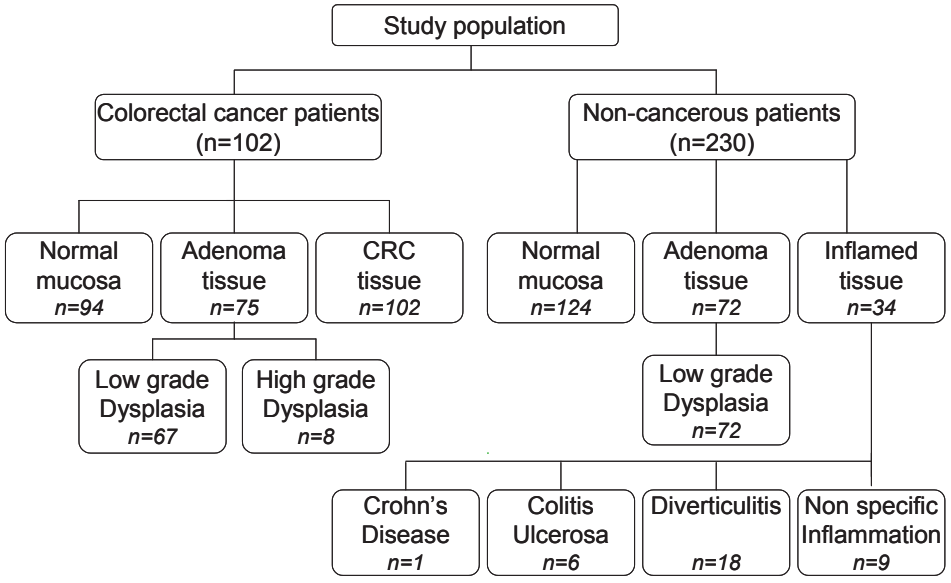

Figure S2.1 Study population.
Of the 102 colorectal cancer patients, carcinoma tissue was collected and, if present, also normal colon mucosa tissue (n=94) and synchronous or metachronous adenoma tissue (n=75). The tissue collected from 230 non-cancerous patients consisted of histologically normal biopsy material (n=124) from patients that underwent endoscopy and did not develop adenomas or colorectal cancer, adenoma biopsies (n=72) from patients who did not develop colorectal cancer within 10 years, and resected colon mucosa from patients with various inflammatory bowel conditions (n=34) who did not develop adenomas or colorectal cancer. CRC: colorectal cancer.

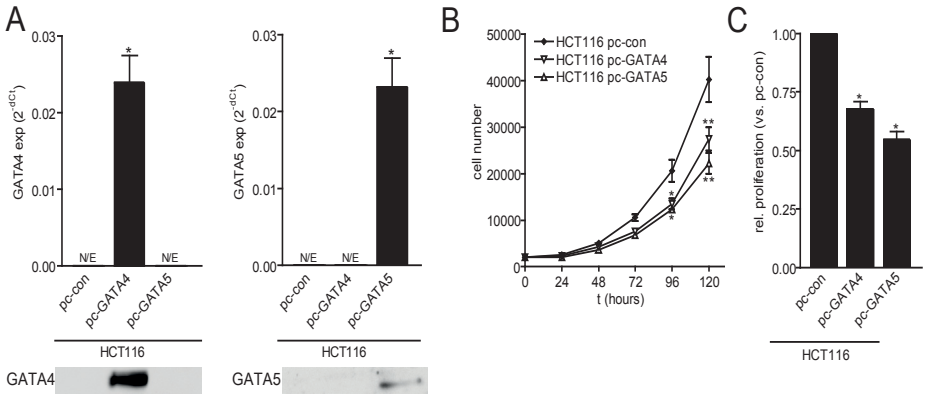

Figure S2.2 **GATA4 and GATA5 inhibit proliferation of HCT116 human colorectal cancer cells.**
A. Bar graphs: GATA4 and GATA5 mRNA expression in HCT116 cells transiently transfected with empty pcDNA3 vector (pc-con), pcDNA3-GATA3-GATA4 vector (pc-GATA4) or pcDNA3-GATA5 vector (pc-GATA5) and selected for 10 days with G418. Results are plotted as mean values (±SEM) of mRNA expression (calculated as 2-dCt) of three independent experiments (*p<0.01 vs. pc-con). N/E not ecpressed. Gel images: Western Blot analysis of GATA4 and GATA5 protein in nuclear extracts of HCT116 cells. **B**. Cell growth of HCT116 cells. Results are plotted as mean values (±SEM) of cell numbers of three independent experiments (*p<0.05 vs. pc-con. **p<0.001 vs. pc-con). **C**. Proliferation measured by 3H-thymidine incorporation. Data are expressed as mean relative proliferation values (±SEM) of three independent triplicate experiments (*p<0.0001 vs. pc-con).

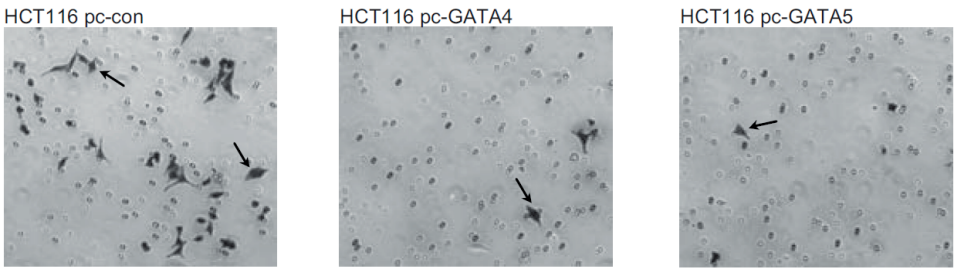

Figure S2.3 **GATA4 and GATA5 decrease invasion of HCT116 human colorectal cells.**
Representative photomicrographics showing matrigel invasion of HCT116 cells transfected with empty pcDNA3 vector (pc-con), pcDNA3 GATA4 vector (pc-GATA4) and pcDNA3 GATA5 vector (pc-GATA5). Examples of trespassed cells are indicated by arrows.

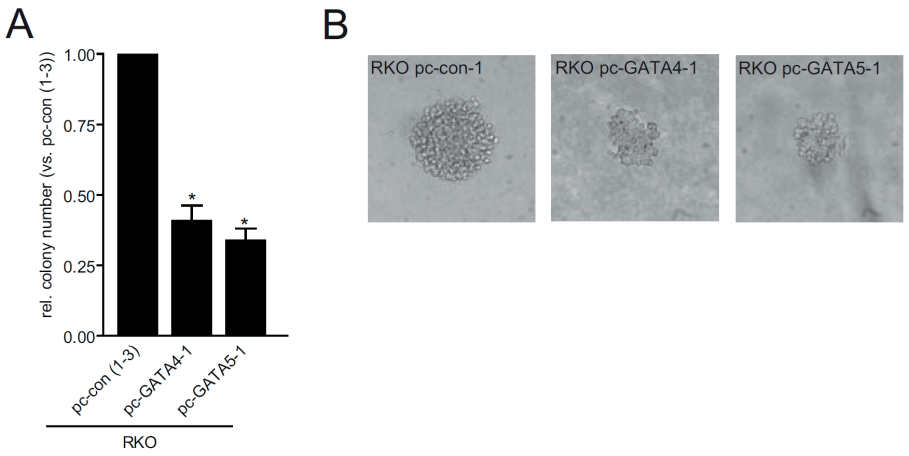

Figure S2.4 GATA4 and GATA 5 decrease anchorage independent growth of RKO human colorectal cancer cells.

A. Anchorage-independent growth of RKO clones as assessed by soft agar colony formation. Data are presented as mean relative numbers of colonies (±SEM) compared with pc-con (1-3) of three independent triplicate experiments (*<0.0001 vs. pc-con (1-3)). **B**. Representative photographs showing colonies of RKO control clone pc-con-1, and RKO clones constitutively expressing GATA4 (pc-GATA4-1) or GATA5 (pc-GATA5-1).

Supplemental tables

Table S2.1 Patient characteristics.

				Location†	
	N	Age*	Sex#	proximal	distal
Colorectal cancer	102				
Normal tissue	94	70.9 ± 8,6	48/46	39/90 (43%)	51/90 (57%)
Adenoma tissue	75	72.0 ± 7.6	40/35	23/72 (32%)	49/72 (68%)
Carcinoma tissue	102	71.4 ± 8.4	48/54	47/100 (47%)	53/100 (53%)
Controls	230				
Normal tissue	124	65.7 ± 9.4	51/73	19/98 (19%)	79/98 (81%)
Adenoma tissue	72	62.8 ± 7.1	43/29	5/60 (8%)	55/60 (92%)
Inflamed tissue	34	65.2 ± 10.0	15/19	4/27 (15%)	23/27 (85%)
P-value		<0.001	ns	<0.001	

			Adenoma tissue	
	Carcinoma tissue		Colorectal cancer	Controls
Histological type		**Histological type**		
Adenocarcinoma	82/102 (80%)	Tubular	48/75 (64%)	57/72 (79%)
Mucinous	20/102 (20%)	Tubulovillous	25/75 (33%)	15/72 (21%)
		Villous	2/75 (3%)	0/72 (0%)
Differentiation				
Poor	10/102 (10%)	**Dysplasia**		
Moderate	80/102 (78%)	lowgrade	67/75 (89%)	72/72 (100%)
Well	12/102 (12%)	Highgrade	8/75 (11%)	0/0 (0%)
TNM stage				
I	16/102 (16%)			
II	34/102 (33%)			
III	40/102 (39%)			
IV	12/102 (12%)			

Of the 102 colorectal cancer patients, carcinoma tissue was collected and, if present, also normal colon mucosa tissue (n=94) and synchronous or metachronous adenoma tissue (n=75). * years ± SD, analyzed by Kruskal-Wallis test; #male/female, analyzed by Pearson's χ^2; †analyzed by Pearson's χ^2. Information of location was not available for all samples, explaining different total sample numbers; TNM stage: "Tumor Node Metastasis" staging; ns: non significant.

Table S22 **Primer sequences.**

Gene	Primer	Sense primer	Antisense primer	Amplicon size (BP)	Annealing temp (°C)	No. of PCR cycles	Ref.
GATA4	flank	5'-GGG AGT TTT TYG TAT AGT TTY GTA G-3'	5'-CCR ACC RCC TCC AAA TCC CCA AC-3'	167	56	35	1
GATA4	U	5'-TTT GTA TAG TTT TGT AGT TTG TGT TTA GT-3'	5'-CCC AAC TCA CAA CTC AAA TCC CCA-3'	140	64	30	1
GATA4	M	5'-GTA TAG TTT CGT AGT TTG CGT TTA GC-3'	5'-AAC TCG CGA CTC GAA TCC CCG-3'	136	68	25	1
GATA5	flank	5'-TAG ATA YGG AGT TYG TTT TTA GGT TAG-3'	5'-CRA AAC CCR AAC CAA TAC AAC TAA AC-3'	160	56	35	1
GATA5	U	5'-TGG AGT TTG TTT TTA GGT TAG TTT TTG GT-3'	5'-CAA ACC AAT ACA ACT AAA CAA ACA AAC CA-3'	147	64	25	1
GATA5	M	5'-AGT TCG TTT TTA GGT TAG TTT TCG GC-3'	5'-CCA ATA CAA CTA AAC GAA CGA ACC G-3'	140	64	25	1
APC-1A	Flank	5'-TGG GYG GGG TTT TGT GTT TTA TT-3'	5'-TAC RCC CAC ACC CAA CCA ATC-3'	136	56	35	2
APC-1A	U	5'-GTG TTT TAT TGT GGA GTG TGG GTT-3'	5'-CCA ATC AAC AAA CTC CCA ACA A-3'	108	60	25	2
APC-1A	M	5'-TAT TGC GGA GTG CGG GTC-3'	5'-TCG ACG AAC TCC CGA CGA-3'	98	60	25	2
$p14^{ARF}$	flank	5'-GYG TTG TTT ATT TTT GGT GTT AAA GG-3'	5'-AAA TAT AAA CCA CRA AAA CCC TCA CT-3'	152	56	35	2
$p14^{ARF}$	U	5'-TTT TTG GTG TTA AAG GGT GGT GTA GT-3'	5'-CAC AAA AAC CCT CAC TCA CAA CAA-3'	132	60	25	2
$p14^{ARF}$	M	5'-GTG TTA AAG GGC GGC GTA GC-3'	5'-AAA ACC CTC ACT CGC GAC GA-3'	122	60	25	2
O^6-MGMT	flank	5'-GYG TTT YGG ATA TGT TGG GAT AGT T-3'	5'-AAA CTC CRC ACT CTT CCR AAA AC-3'	135	56	35	2
O^6-MGMT	U	5'-TTT GTG TTT TGA TGT TTG TAG GTT TTT GT-3'	5'-AAC TCC ACA CTC TTC CAA AAA CAA AAC A-3'	93	60	30	2
O^6-MGMT	M	5'-TTT CGA CGT TCG TAG GTT TTC GC-3'	5'-GVA CTC TTC CGA AAA CGA AAC G-3'	81	60	30	2
HLTF	flank	5'-GTA GGT ATY GTA GTY GTA TTT TTG GG-3'	5'-CAA AAC ACA AAA AAA AAA ACA ACT CC-3'	155	56	35	3
HLTF	U	5'-GGT TTT GTG GTT TTT TTG TGT GTT T-3'	5'-CCC CAC TAC CAT TCA AAA ACA ACA-3'	103	65	30	3
HLTF	M	5'-GTG GTT TTT TCG CGC GTT C-3'	5'-CGC TAC CAT TCA AAA ACG ACG-3'	96	65	30	3
$p16^{INK4A}$	flank	5'-GGG TTG GTT GGT TAT TAG AGG GT-3'	5'-RAC CRT AAC CAA CCA ATC AAC C-3'	148	56	35	2
$p16^{INK4A}$	U	5'-GTT GGT TAT TAG AGG GTG GGG TGG ATT GT-3'	5'-AAC CAA AAA CTC CAT ACT ACT CCC CAC CA-3'	124	62	25	2
$p16^{INK4A}$	M	5'-TTA TTA GAG GGT GGG GCG GAT CGC-3'	5'-GAA AAC TCC ATA CTA CTC CCC GCC G-3'	115	62	25	2
RASSF1A	flank	5'-GTT TAG TTT GGA TTT TGG GGG AG-3'	5'-CCC RCA ACT CAA TAA ACT CAA ACT C-3'	144	56	35	2
RASSF1A	U	5'-GGG GTT TGT TTT GTG GTT TTG TTT-3'	5'-AAC ATA ACC CAA TTA AAC CCA TAC TTC A-3'	81	64	30	2
RASSF1A	M	5'-GGG TTC GTT TTG TGG TTT CGT TC-3'	5'-TAA CCC GAT TAA ACC CGT ACT TCG-3'	76	64	30	2

Primers and PCR conditions for nested MSP.

ref 1: Akiyama et al. Mol Cell Biol 2003;23:8429-39, ref 2: van Engeland et al. Cancer Res 2003;63:3133-7, ref 3: Brandes et al. Carcinogenesis 2005;26:1152-6

Gene	Sense primer	Antisense primer
Cyclophilin A	5'-CTC GAA TAA GTT TGA CTT GTG TTT-3'	5'-CTA GGC ATC GGA GGG AAX A-3'
GATA4	5'-CTG GCC TGT CAT CTC ACT ACG-3'	5'-GGT CCG TGC AGG AAT TTG AGG-3'
GATA5	5'-CCT GCG GCC TCT ACC ACA A-3'	5'-GGC GCG GCG GGA CGA GGA C-3'

Primer for real-time RT-PCR

Beta-actin	5'-TAG GGA GTA TAT AGG TTG GGG AAG TT-3'	5'-AAC ACA CAA TAA CAA ACA CAA ATT CAC-3'	5'-FAM-CGA CTG CGT GTG GGG TGG TGA TGG AGG AGG TTT AGG CAG TCG-3'-DABCYL
GATA4	5'-AGG TTA GTT AGC GTT TTA GGG TC-3'	5'-ACG ACG ACG AAA CCT CTC G-3'	5'-FAM-CGA CAT GC TCG CGA CTC GAA TCC CCG ACC CAG CAT GTC G-3'-DABCYL

Primers for quantitative MSP

Table S2.3 Subjects characteristics (stool samples).

			Training set	Test set
Mean age	controls		55	52
	CRC patients		69	71
Sex	controls	Female	26/45	22/30
		Male	19/45	8/30
		Unknown	0/45	0/30
	CRC patients	Female	8/28	19/47
		Male	19/28	27/47
		Unknown	1/28	1/47
Stage	I		10/28	10/47
	II		8/28	19/47
	III		8/28	13/47
	IV		2/28	4/47
	unknown		0/28	1/47

CRC: colorectal cancer

Table S2.4 GATA4/5 methylation frequencies in adenoma and normal tissue from patients with and without colorectal cancer.

	Adenoma tissue			Normal tissue		
	CRC-	CRC+	P-value* (<)	CRC-	CRC+	P-value* (<)
GATA4	17/52 (33%)	28/56 (50%)	ns	5/88 (6%)	8/81 (10%)	ns
GATA5	23/56 (41%)	29/47 (62%)	ns	13/100 (13%)	8/66 (12%)	ns

Logistic regression was used to adjust for age and location. * Bonferroni-corrected p-value. ns: not significant. CRC+: colorectal cancer patients, CRC-: non-cancerous individuals.

Table S2.5 GATA4/5 methylation frequencies in carcinoma, adenoma and normal tissue from colorectal cancer patients.

	CRC+								
	Normal	Adenoma	P-value* (<)	Normal	Adenoma	P-value* (<)	Normal	Adenoma	P-value* (<)
GATA4	4/44 (9%)	21/44 (48%)	$2^{*}10^{-3}$	4/53 (8%)	36/53 (68%)	$4^{*}10^{-7}$	20/43 (47%)	31/43 (72%)	ns
GATA5	6/34 (15%)	31/34 (62%)	$5^{*}10^{-4}$	6/39 (15%)	31/39 (80%)	$2^{*}10^{-7}$	23/36 (64%)	28/36 (78%)	ns

Differences in methylation percentages between normal and adenoma tissue, between normal and carcinoma tissue, and between adenoma and carcinoma tissue from 61 colorectal cancer patients for which all three tissues could be collected, analyzed by the Mc Nemar test. Frequencies may vary because of missing data leading to differences in numbers of patients with simultaneous information on two types of tissues (e.g. not all patients with adenomas also had both normal and carcinoma tissue available). * Bonferroni-corrected p-value. ns: not significant. CRC+: colorectal cancer patients, CRC-: non-cancerous individuals.

Chapter 3

The emerging role of GATA transcription factors in development and disease

Marjolein H.F.M. Lentjes, Hanneke E.C, Niessen, Yoshimitsu Akiyama,
Adriaan P. de Bruïne, Veerle Melotte, Manon van Engeland

Expert Reviews in Molecular Medicine 2016;18:e3

Abstract

The GATA family of transcription factors consists of six proteins (GATA1-6) which are involved in a variety of physiological and pathological processes. GATA1/2/3 are required for differentiation of mesoderm and ectoderm-derived tissues, including the haematopoietic and central nervous system. GATA4/5/6 are implicated in development and differentiation of endoderm- and mesoderm-derived tissues such as induction of differentiation of embryonic stem cells, cardiovascular embryogenesis and guidance of epithelial cell differentiation in the adult.

Introduction

The importance of GATA factors for development is illustrated by the embryonic lethality of most single *GATA* knockout mice. Moreover, *GATA* gene mutations have been described in relation to several human diseases, such as hypoparathyroidism, sensorineural deafness and renal insufficiency (HDR) syndrome, congenital heart diseases (CHDs) and cancer. GATA family members are emerging as potential biomarkers, for instance for the risk prediction of developing acute megalokaryblastic leakemia in Down syndrome and for the detection of colorectal- and breast cancer.

The origin and molecular structure of the GATA family

In vertebrates, six GATA transcription factors have been identified. Based on phylogenetic analysis and tissue expression profiles, the GATA family can be divided into two subfamilies, GATA1/2/3 and GATA4/5/6.[1] Although in non-vertebrates *GATA* genes are linked together onto chromosomes, in humans they are segregated onto six distinct chromosomal regions (Table 3.1), indicating segregation during evolution.[2] Most *GATA* genes encode for several transcripts and protein isoforms. GATA proteins have two zinc finger DNA binding domains, Cys-X2-C-X17-Cys-X2-Cys (ZNI and ZNII), which recognise the sequences (A/T)GATA(A/G) (Figure 3.1).[3] Amongst the six GATA binding proteins, the zinc finger domains are more than 70% conserved, while the sequences of the amino-terminal and carboxyl-terminal domains exhibit lower similarity.[4] In non-vertebrates GATA transcription factors have been identified that contain mostly one zinc finger, i.e. in *Drosophila melanogaster* and *Caenorhabditis elegans*.[3] The C-terminal zinc finger (ZNII) exists in both vertebrates and non-vertebrates indicating that ZNI was duplicated from ZNII.[2]

Tissue-specific roles of GATA factors in development and disease

Haematopoietic system

GATA1/2/3 knockout mice die at the embryonic stage due to haematological abnormalities (Table 3.2), indicating a pivotal role of these transcription factors in haematopoietic development.[1]

GATA1, the first recognised member of the GATA family, is specifically expressed during haematopoietic development of erythroid, and megakaryocytic cell lineages (Figure 3.2).[11] Loss of *GATA1* in mouse embryo-derived stem cells results in a complete lack of primitive erythroid precursor production.[5] Definitive erythroid precursors, on the other hand, are normally produced, but undergo a maturation arrest at the proerythroblast stage followed by apoptosis.[12] Ablation of *GATA1* in adult mice also results in a maturation arrest at the same proerythroblast stage.[13]

Table 3.1 Molecular features of the human GATA transcription factors.

	Genomic sequence			mRNA sequence			Protein sequence		
Name	Chromosomal location	Ensembl Acession No.	CpG island in the promoter region	Transcripts	Ensembl Transcript ID *	coding exons	Uniprot Accession No.	Isoform	protein Length (AA)
GATA1	Xp11.23	ENSG00000102145	None	3	ENST00000376670	5	P15976	1	413
								2	330
								3	335
GATA2	3q21.3	ENSG00000179348	+	6	ENST00000341105	5	P23769	1	480
								2	466
GATA3	10p15	ENSG00000107485	+	5	ENST00000346208	5	P23771	1	443
								2	444
GATA4	8p23.1-p22	ENSG00000136574	++	9	ENST00000335135	6	P43694		442
GATA5	20q13.33	ENSG00000130700	++	4	ENST00000252997	6	Q9BWX5		397
GATA6	18q11.1-q11.2	ENSG00000141448	++	1	ENST00000269216	6	Q92908	1	595
								2	449

* In the case of multiple transcripts the Ensembl Transcript ID was chosen, based on the first isoform of the corresponding Uniprot protein sequence.

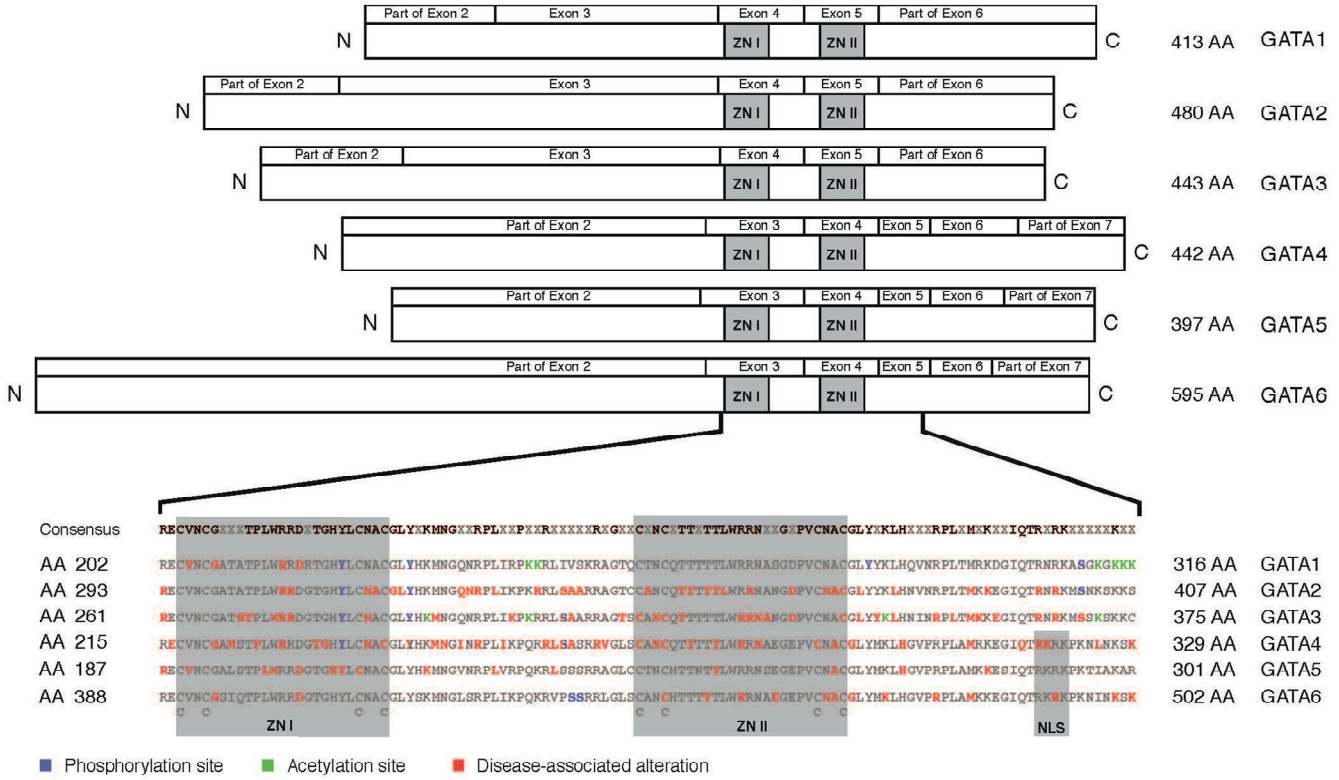

Figure 3.1 **Overview of GATA1-6 proteins.** The GATA proteins are depicted in the upper part of the figure. The GATA proteins are aligned according to the location of the zinc fingers (ZNI and ZNII). The exon boundaries are depicted above the protein structure. For GATA4 the transcriptional activation domains (TADI and TADII) are shown. In the lower part of the figure the regions around the zinc fingers are enlarged, with the corresponding amino acid (AA) numbers written next to the GATA sequence. Posttranslational modification (post-transcriptional modification) sites and disease-associated alterations are marked on top of the corresponding AA.

The requirement of the different GATA1 functional domains during primitive and definitive erythropoiesis has been investigated in vivo, showing that both zinc fingers are needed to rescue *GATA1* germline mutant mice.[14] In haematopoietic stem cells (HSCs), *GATA1* gene expression is suppressed, which is indispensable for the maintenance of these stem cells. The mechanism behind this suppression is not fully understood yet. Recently, it was shown that decreased DNA methylation of the *GATA1* locus leads to increased GATA2 binding and that increased GATA2 binding results in *GATA1* gene transactivation. According to these study results, Takai et al. proposed a mechanism in which *GATA1* hypomethylation results in an accessible locus for GATA2 binding which enables transactivation of *GATA1* gene expression to initiate erythropoiesis in megakaryo-erythroid progenitors.[15] Loss of *GATA1* results in a marked increase of GATA2 expression, indicating not only that GATA2 partially compensates for GATA1 but also that GATA1 suppresses GATA2 transcription during normal erythropoiesis.[16] This suppression is mediated by the displacement of GATA2 from its upstream enhancer by increasing levels of GATA1 referred to as the 'GATA switch'.[17] The combined loss of *GATA1* and *GATA2* in double-knockout embryos leads to an almost complete absence of primitive erythroid cells, suggesting functional overlap between these transcription factors early in the primitive erythropoiesis.[18]

Table 3.2 Phenotype of GATA knockout mice.

Name	Phenotype (embryonic day)	Abnormality	Reference
GATA1	die (11.5-12.5 dpc)	defective erythroid cell maturation	5
GATA2	die (12.5 dpc)	severe anemia	6
GATA3	die (11-12 dpc)	massive internal bleeding and severe deformities of the brain and spinal cord	7
GATA4	die (9.5 dpc)	defects of heart morphogenesis and ventral closure of the forgut	8
GATA5	viable and fertile	females exhibited pronounced genitourinary abnormalities that included vaginal and uterine defects and hypospadias	9
GATA6	die (5.5- 7.5 dpc)	defects of visceral endoderm function and subsequent extra-embryonic endoderm	10

Dpc, days post coïtum

Requirement of functional GATA1 for haematopoiesis is also observed in several human diseases, such as anaemia, leukaemia and thrombocytopenia (Table 3.3). Splice site mutations of *GATA1* have been found in a family with macrocytic anaemia and in patients with Diamond-Blackfan anaemia (an anaemia characterised by a selective hypoplasia of erythroid cells), resulting in impaired production of the full-length form of the GATA1 protein.[19,20]

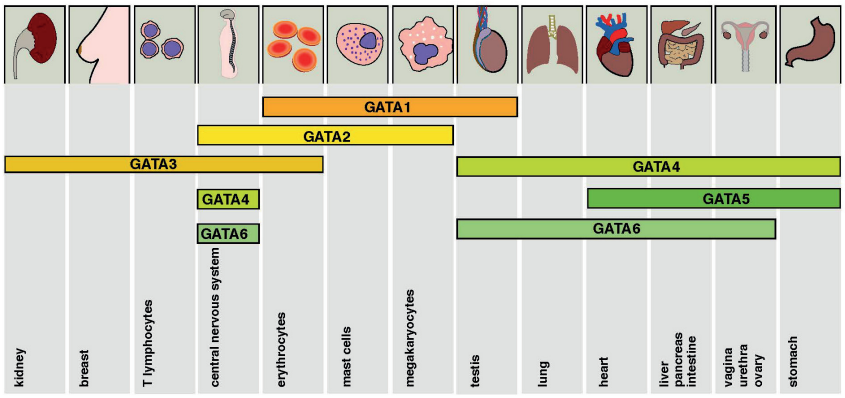

Figure 3.2 **Distribution of GATA expression in various organs during vertebrate development**.
The expression of all GATA factors is depicted in the corresponding tissues. The distribution of the expression patterns roughly reflects the two GATA subgroups (GATA1/2/3 versus GATA4/5/6).

Conditional megakaryocytic lineage specific *GATA1* knockout mice show excessive marrow megakaryocyte proliferation whereas the platelet numbers are decreased. The maturation of these hyperproliferated megakaryocytes is severely impaired and the produced platelets are structurally and functionally abnormal.[21] Additionally, megakaryocyte-expressed genes with functional GATA1-binding sites (e.g. *STAT1*) are downregulated in *GATA1*$^{-/-}$ megakaryocytes.[22] Loss of GATA1 leads to overexpression of GATA2 in megakaryocytes. However GATA1-deficient megakaryocytes still show abnormal megakaryocytic proliferation and differentiation, establishing no functional redundancy of these transcription factors in megakaryopoiesis.[23] In contrast to erythropoiesis, GATA2 remains to be expressed after the GATA switch in late megakaryopoiesis, suggesting a divergent function for both GATA proteins.[24]

Children with trisomy 21 are at risk of developing leukaemia, in particular acute megakaryoblastic leukaemia (AMKL). Nearly all Down syndrome patients with AMKL harbour somatic mutations in the *GATA1* gene (Table 3.3)[25], predominantly leading to an N-terminal truncated 'short' GATA1 protein (GATA1s).[26] Inadequate GATA1 mediated repression of specific oncogenic factors contributes to megakaryocytic abnormalities.[27] Analysis of Down syndrome children with transient myeloproliferative disorder (TMD), which is considered a potential precursor to AMKL, also revealed *GATA1* mutations.[28] Noticeable the *GATA1* mutation in TMD and subsequent AMKL is identical, suggesting that *GATA1* mutations are early events in the development of AMKL in trisomy 21-children.[29] Not all TMD Down syndrome neonates with a *GATA1* mutation progress to AMKL, indicating the need for more molecular events contributing to the pathogenesis of AMKL. Recently, Yoshida et al. reported newly acquired driver mutations, which lead to the development from TMD to Down syndrome-AMKL.[30,31]

Table 3.3 GATA transcription factors in disease.

	Disease	Abberation	Location	Consequence
GATA1	XLT	MS mut	ZnF1	FOG1 interaction ↓
	XLTT	MS mut	ZnF1	DNA binding ↓
	Anemia (e.g. Diamond-Blackfan anemia)	Splice site mutation, mutation initiation codon	exon 2	only short for or los of the full length GATA1 isoform
	Congenital erythropoietic porphyria	MS mut	ZnF1	unknown
	TMD and AMKL in DS	FS INS and DEL, NS mut and splice site mutation	Intron 1, exon 2 and 3	protein truncation, transcriptional activation ↓
	AMKL without DS	INS	exon 2	protein truncation
GATA2	Chronic myeloid leukemia	MS mut, FS DEL	ZnF2	DNA binding ↑, transcriptional activation ↓
	DCML / MonoMAC / Emberger syndrome	FS INS and DEL, MS and NS mut, full gene DEL	ZnF2, 5'UTR, intron 5	nonfunctional protein, nonsense-mediated decay
	Myelodysplastic syndrome	FS INS and DEL, MS and NS mut, full gene DEL	exon, intron, 5'UTR	protein truncation, DNA binding ↓
	Acute myeloid leukemia	MS mut, FS INS, full gene DEL	ZnF1, ZnF2, exons	nonfunctional protein
GATA3	HDR syndrome	MS and NS mut, FS INS and DEL, splice site mutation,partial and full gene DEL	ZnF1, ZnF2, exons	protein truncation, FOG2 interaction ↓, DNA binding/affinity ↓
	Breast cancer	MS and NS mut, FS INS and DEL	ZnF2, exons	protein truncation, nonfunctional protein
	T-ALL	MS mut, FS DEL, in-frame DEL	ZnF1, ZnF2	likely loss of function
	B-ALL	SNP	intron 3	unknown
	UCC and RCC	CpG methylation	promoter	transcriptional activation ↓
GATA4	CHD	MS and NS mut, FS INS and DEL, SNP, full gene DEL, gene duplication	ZnF1, ZnF2, exons, 3´UTR, introns, promoter	protein truncation, DNA binding/affinity ↓, transcriptional activation ↓, TBX5 interaction ↓, changed RNA folding
	Pancreatic agenesis	MS mut, intragenic and full gene deletion	ZnF2	transcriptional activation ↓, DNA binding ↓
	GI cancer	CpG methylation, amplification	promoter, 8p	transcriptional activation ↓/↑
	Glioblastoma multiforme	CpG methylation, FS INS and DEL	promoter, ZnF domains, C terminal region	transcriptional activation ↓
	Ovarian cancer	hypoacetylation, loss trimethylation, CpG methylation	histone 3 and 4, lysine 4	transcriptional activation ↓
	Other cancers (e.g. lung, DLBCL)	CpG methylation	promoter	transcriptional activation ↓

Table 3.3 (continued)

	Disease	Abberation	Location	Consequence
GATA5	CHD	MS and NS mut	ZnF1, ZnF2	transcriptional activation ↓
	Cancer (e.g. GI cancer, RCC)	CpG methylation	promoter	transcriptional activation ↓
GATA6	CHD	MS and NS mut, duplication and DEL	ZnF1, ZnF2, exons	transcriptional activation ↓
	Pancreatic agenesis	MS and NS mut, FS INS and DEL	Znf2, exons	transcriptional activation ↓
	Ovarian cancer	hypoacetylation, loss trimethylation, upregulation	histone 3 and 4, lysine 4	transcriptional activation ↓/↑
	GI cancer	amplification, CpG methylation	18q, promoter	transcriptional activation ↓/↑
	Pancreatobiliary cancer	amplification	18q11.2	transcriptional activation ↑
	Pediatric rhabdomyosarcoma	CpG methylation	promoter	transcriptional activation ↓

AMK acute megakaryoblastic leukemia; B-ALL B-cell acute lymphoblastic leukemia; CHD congenital heart disease; CML chronic myeloid leukemia; DCML dendritic cell, monocyte, B-lymphocyte and natural killer lymphocyte deficiency; DEL deletion; DLBCL diffuse large B-cell lymphoma; DS Down syndrome; GI cancer gastrointestinal cance; FS frameshift; HDR hypoparathyreoidism, sensorineural deafness and renal disease; INS insertion; MS mut missense mutation; MonoMAC syndrome associated with monocytopenia, B and NK cell lymphopenia and mycobacterial, fungal and viral infections; NS mut nonsense mutation; RCC renal cell carcinoma; SNP single nucleotide polymorphism; T-ALL T-cell acute lymphoblastic leukemia; TMD transient myeoloproliferative disorder; UCC urothelial cell carcinoma; XLT X-linked thrombocytopenia; XLTT X-linked thrombocytopenia with thalassemia.

The mechanism behind the leukaemogenesis remains elusive. Based on mutational spectrum analysis of the *GATA1* locus in Down syndrome AMKL, Cabelof et al. hypothesised that increased oxidative stress because of trisomy 21, uracil accumulation and reduced DNA repair together driving leukaemogenesis in Down syndrome.[32] Recently it was shown that *GATA1* mutations protect megakaryocytes from activated AKT-induced apoptosis.[33] Additionally, trisomy 21 itself increases HSC frequency, clonogenicity and megakaryocyte-erythroid output with associated megakaryocyte-erythroid progenitor expansion.[34,35,36] Another hypothesis is that upregulation of runt-related transcription factor 1 (RUNX1), which physically interacts with GATA1, due to trisomy 21 leads to the induction of GATA1 transcription during embryogenesis, thereby leading to transcription-associated mutagenesis.[37] Recently it is shown that loss of type I interferon (IFN) signalling contributes to GATA1s-induced megakaryocyte hyperproliferation, suggesting AMKL treatment with IFN-α administration.[38]

GATA1 mutations are also detected in a specific form of X-linked hereditary thrombocytopenia and are described with and without thalassemia (Table 3.3 and Supplemental Table S3.1). Hereditary thrombocytopenia without thalassemia has been associated with *GATA1* missense mutations that are located in the N-terminal zinc finger region. These mutations lead to loss or inhibition of GATA1 interaction with friend-of-GATA(FOG)1-cofactor.[39] The degree of disrupted GATA1–FOG1 interaction depends on the mutation, explaining different clinical presentations.[40] The only *GATA1* mutation reported in hereditary X-linked thrombocytopenia with thalassemia is the missense mutation R216Q which is located in the DNA binding surface of the GATA1 N-terminal zinc finger and results in reduced DNA binding rather than affecting GATA1-FOG1 interaction.[41]

In vertebrates, *GATA2* is expressed in haematopoietic progenitor cells (HPCs), early erythroid cells, mast cells and megakaryocytes, closely resembling the cellular distribution of GATA1 (Figure 3.2). A deficit in primitive erythropoiesis is apparent in *GATA2*$^{-/-}$mice since the total number of blood cells during embryonic development is markedly reduced, leading to lethality because of severe anaemia (Table 3.2).[6] In *GATA2*$^{+/-}$ mice haematopoietic defects are seen within HSCs and granulocyte-macrophage progenitor cells. Moreover, the loss of *GATA2* in adult mice leads to profound abnormalities in definitive haematopoiesis, also directing to a defect at the level of HSCs.[6,42,43] The function of GATA2 in haematopoietic development has recently been reviewed by Bresnick et al.[44], describing GATA2 as one of the key components establishing the transcriptional program for early haematopoietic development.

Two different *GATA2* alterations have been reported in patients with chronic myeloid leukemia (CML) during blast crisis formation (Table 3.3). In contrast to the in-frame deletion Δ341-346, which leads to decreased transcriptional activation, *GATA2* L359V is a gain-of-function mutation and leads to increased DNA binding. Transduction of *GATA2* L359V (in vitro and in vivo) resulted in disturbed myelomonocytic

differentiation/proliferation, suggesting *GATA2* mutations are involved in the acute myeloid transformation of CML.[45]

GATA2 gene mutations that predisposed to myelodysplastic syndrome (MDS) and acute myeloid leukaemia (AML) were reported (Supplemental Table S3.1). This occurred either in the absence (non-syndromic) or presence of certain syndromes, including Emberger syndrome and monoMAC syndrome.[46] Most mutations affect the C-terminal zinc finger or result in N-terminal frameshift mutations.[47]

Similar expression patterns of GATA1, GATA2 and GATA3 in human, murine and avian erythroid cells indicate a conserved role for these GATA transcription factors in vertebrate erythropoiesis.[48] Beyond its expression in erythroid lineages, GATA3 is also expressed in T lymphocytes.[49] During haematopoiesis vertebrate GATA3 is expressed in HSCs and in developing T lymphocytes. Murine *GATA3*$^{-/-}$ embryos are predominantly affected during definitive haematopoiesis in the fetal liver. Although later than *GATA2*$^{-/-}$ mice, these embryos appear also anaemic and die in utero, probably owing to massive internal bleeding (Table 3.2).[7] Frelin et al. demonstrated that GATA3 regulates the self-renewal and differentiation of bone marrow long-term HSCs.[50] During embryogenesis, GATA3 deficiency leads to a marked reduction in the production of HSCs in the aorta-gonads-mesonephros region. It was shown that GATA3 regulates HSC emergence during embryogenesis via the production of catecholamines linking the haematopoietic system development to the development of the sympathetic nervous system (SNS).[51]

In T cell development, GATA3 has a pivotal role from the generation of early T lineage progenitors to CD4+ specification [as reviewed by Hosoya T et al.[52]]. During antigen presentation by specialised antigen-presenting cells, the TCR is stimulated, thereby driving differentiation from peripheral naïve $CD4^+$ T cells towards T helper cell type 1(T_H1) or 2 (T_H2). GATA3 expression in differentiating T_H2 cells is mediated by different pathways as clearly reviewed in Ho et al..[53] GATA3 and STAT6 in T_H2 lineage account for lineage specific expression of T cell lincRNAs. At the moment, the function of lincRNAs during T cell development and differentiation is under investigation.[54] An essential function for GATA3 beyond T_H2 differentiation is also described demonstrating GATA3 controls proliferation and maintenance of mature T cells.[55]

GATA3 dysregulation is described in leukaemia. Together with T-cell acute lymphocytic leukemia 1 (TAL1) and RUNX1, GATA3 forms an autoregulatory loop that positively regulates the v-myb avian myeloblastosis viral oncogene (MYB) oncogene, which in turn controls the gene expression program in T-cell acute lymphoblastic leukaemia (T-ALL).[56] Thereby, whole-genome sequencing of patients with early T-cell precursor ALL, an aggressive subtype of T-ALL, revealed *GATA3* inactivating mutations (Supplemental Table S3.1).[57]

In summary, GATA1/2/3 are essential regulators in the development of erythroid and megakaryocytic cell lineages and in the molecular pathogenesis of different haematopoietic diseases.

Cardiovascular system

The mesoderm gives rise to numerous organs, including the heart and genitourinary tract. GATA4/5/6 proteins are expressed in the mesodermal precursors that develop into the heart.[58]

GATA4 is one of the earliest transcription factors expressed in developing cardiac cells, already detectable in murine precardiac splanchnic mesoderm and associated endoderm.[8] *GATA4*$^{-/-}$ mice display severe defects in ventral foregut closure and heart morphogenesis, resulting in embryonic lethality at embryonic day 8 (Table 3.2). These deformities result from a general loss in ventral folding throughout the embryo and implicate GATA4 requirement for the migration or folding morphogenesis of the precardiogenic splanchic mesodermal cells.[8] Mice harbouring a knock-in mutation that abrogates the interaction with FOG-cofactors (*GATA4*$^{Ki/Ki}$) lack coronary vessels.[59] In addition, murine GATA4 regulates cardiac angiogenesis by inducing angiogenic factors such as VEGF, facilitating compensation following injury.[60] Yamak et al. have suggested that GATA4 and Cyclin D2 are part of a forward reinforcing loop in which Cyclin D2 feeds back to enhance cardiogenic activity of GATA4 through direct interaction. *GATA4* mutations that abrogate Cyclin D2 interactions are associated with human CHD.[61]

A variety of *GATA4* mutations have been detected in patients with various forms of CHD such as Tetralogy of Fallot, ventricular septal defect and atrial fibrillation as reviewed by McCulley et al. and summarised in Table 3.3 and Supplemental Table 3.1.[62]

Within the developing heart, GATA5 is expressed in the myocardium as well as in the endocardium and derived endocardial cushions in mouse embryos.[63] Depending on how *GATA5* is inactivated in several mouse models, different cardiac phenotypes are described. Deletion of both *GATA5* isoforms leads to hypoplastic hearts and partially penetrant bicuspid aortic valve formation.[64] When a *GATA5* mutant allele was established that lacked the two zinc finger domains, cardiovascular defects were only detectable in a *GATA4*$^{+/-}$ background.[65] Although little is known about GATA5 in human heart conditions, three heterozygous *GATA5* mutations have been associated with familial atrial fibrillation[66] and four heterozygous *GATA5* mutations with CHD.[67]

GATA6 is abundantly expressed in vascular smooth muscle cells during murine embryonic and postnatal development.[68] *GATA6*$^{-/-}$mice die at the embryonic stage due to defects of the extra-embryonic endoderm (Table 3.2).[10] Tissue-specific deletion of *GATA6* in neural crest-derived smooth muscle cells results in an interrupted aortic arch and persistent truncus arteriosus (PTA). These results suggest that GATA6 is required for proper patterning of the aortic arch arteries. This phenotype is associated with severely attenuated expression of semaphorin 3C, a signalling molecule critical for both neuronal and vascular patterning.[69] Other GATA6 target genes, e.g. *Wnt2*, in vascular smooth muscle cells and cardiac cells have been identified by microarray analysis after transient GATA6 over-expression. Interestingly, *GATA6* is also a target of Wnt2 and together they form a feedforward transcriptional loop to regulate posterior cardiac development.[70]

A number of mutations have been described for *GATA6* in the aetiology of CHD (Table 3.3; Supplemental Table S3.1). For example, two *GATA6* mutations were found in patients with PTA disrupting the transcriptional activity of the GATA6 protein on downstream genes involved in the development of the cardiac outflow tract.[71]

Thus, the GATA4/5/6 transcription factors have closely related functions during cardiovascular development, and defects lead to CHD and other heart conditions.

Gastrointestinal tract

The endoderm gives rise to the respiratory and gastrointestinal tract as well as the associated organs such as pancreas and liver. Differentiation of embryonic stem cells towards the extra-embryonic endoderm can be induced by forced expression of either GATA4 or GATA6.[72] Targeted mutagenesis of *GATA4* in mouse embryonic stem cells results in disturbed differentiation of the visceral endoderm, suggesting that GATA4 has a role in yolk sac formation.[73]

Murine GATA4 is expressed in the proximal but not in the distal small intestine and has an important role in the maintenance of jejunal-ileal identities.[74] Furthermore, GATA4 is essential for jejunal functions such as fat and cholesterol absorption.[75] Beuling et al. found that reduction of GATA4 activity in the intestine induces bile acid absorption in the proximal ileum, which can restore bile acid homeostasis in mice with an ileocaecal resection.[76]

Whereas GATA4 expression is absent from the distal ileum, GATA6 is expressed throughout the entire small intestine. Conditional deletion of *GATA6* in the ileum results in a decrease of crypt cell proliferation and numbers of enteroendocrine and Paneth cells, an increase in numbers of goblet-like cells in crypts and altered expression of genes specific to absorptive enterocytes. GATA4/6 factors are therefore required for proliferation, differentiation and gene expression in the small intestine.[77]

In humans, GATA4 and GATA5 are expressed in normal gastric and colon mucosa.[78,79] In gastric and colorectal cancer (CRC) these genes are frequently transcriptionally silenced by methylation.[78,80] In addition, we reported that GATA4 and GATA5 exhibit tumour suppressive properties in human CRC cells in vitro.[80] The potential biomarker capacities of *GATA4* are discussed below.

Liver and pancreas

In the mouse, the ventral foregut endoderm differentiates to form the parenchymal components of the liver and ventral pancreas. Although GATA4 has an essential function in embryonic liver development, the protein seems to be dispensable in the adult liver function.[81,82] *GATA6*$^{-/-}$ murine embryos have defects in endoderm differentiation, and show severely attenuated GATA4 expression levels and complete absence of hepatocyte nuclear factor 4 (HNF4) expression in the visceral endoderm, parietal endoderm and liver bud.[83] HNF4 is a key regulator for complete differentiation of visceral endoderm, hepatocyte differentiation and the epithelial transformation of

the liver.[84] Tetraploid rescue experiments with *GATA6* null mice show that GATA6 is a key regulator for liver bud growth and commitment of the endoderm to a hepatic cell fate.[83]

Development of the ventral pancreas was, in contrast to the dorsal pancreas, impaired in *GATA4*$^{-/-}$ murine embryos using tetraploid rescue experiments. *GATA6*$^{-/-}$ embryos show a similar phenotype, although not as severe as that observed in *GATA4*$^{-/-}$ embryos.[81] In humans, the role of GATA6 in pancreatic development became apparent in a group of patients with pancreatic agenesis, in which Allen et al. identified 15 de novo heterozygous inactivating mutations in *GATA6* (Supplemental Table S3.1). In addition, these patients suffered from CHD, biliary tract abnormalities, gut developmental disorders, neurocognitive abnormalities and other endocrine abnormalities.[85] In contrast to these results, Martinelli et al. described that GATA6 is dispensable for pancreas development. However, GATA6 is essential for acinar differentiation and maintenance of adult exocrine homeostasis in mice.[86] An explanation for this contradiction might be the timepoint of GATA6 inactivation which is earlier in agenesis patients compared with the mouse model used by Martinelli et al. Together these data show the need for further research to unravel the role of GATA6 in pancreatic development.

In pancreatic cancer, GATA6 is often overexpressed, which correlates with *GATA6* amplification (Table 3.3).[87] Retained GATA6 expression has been shown in gastric, colorectal, esophageal, ovarian and pulmonary cancer cell lines.[78,88,89,90] Additionally, intestinal GATA6 expression is higher in proliferating progenitor cells compared with differentiated cells.[91] In primary gastric cancer, the pro-oncogenic effects of GATA6 are recently confirmed, in vitro and in vivo.[92]

Urogenital tract and kidney

GATA1 is abundantly expressed in the Sertoli cells of the testis during murine prepubertal testis development (Figure 3.2). GATA1 expression decreases thereafter and is in the adult mouse testis only found in the Sertoli cells during different stages of the spermatogenesis.[93] Surprisingly, Sertoli-specific *GATA1* knockout mice show no alterations in testis development, spermatogenesis, male fertility and expression of putative testis-specific GATA1 target genes.[94] Further research has to clarify whether there is a functional redundancy between GATA factors in the testis.

During urogenital development, *GATA4* is expressed in somatic ovarian and testicular cell lineages, and is suggested to have an important regulatory role in gonadal gene expression (Figure 3.2).[95] Mouse embryos conditionally deficient in *GATA4* show no formation of the genital ridge, the structure which differentiates into either testis or ovary.[96] *GATA4*$^{ki/ki}$ mice and *FOG2*$^{-/-}$ mice display defects in the gonadogenesis in both sexes.[97] SRY (Y chromosome-linked testis-determining gene), MIS (Mullerian inhibiting substance) and SOX9 expression, which is critical for testis formation, are dependent on GATA4 × FOG2 interaction.[98] Recently, a signalling cascade was suggested describing

transduction of the p38 mitogen-activated protein kinase (MAPK) pathway by MAP3K4 and GADD45G which leads to GATA4 phosphorylation and thereby activation. Phosphorylated GATA4 then binds and activates the *SRY* promoter.[99]

The *GATA4* gene has also been implicated in a disorder of sex development (DSD). A *GATA4* mutation, which abrogates the binding with FOG2, was discovered in a family with both CHD and 46,XY DSD (Table 3.3).[100] The phenotype closely resembles that of the mouse *GATA4*$^{ki/ki}$ model.[97] The data described above indicate that GATA4, in combination with FOG2, is necessary for proper mammalian sex differentiation.

Murine GATA5 is expressed in the urogenital ridge during foetal development.[63] *GATA5*$^{-/-}$ female mice exhibit abnormalities of the genitourinary tract including malpositioning of the urogenital sinus, vagina and urethra, whereas males are unaffected (Table 3.2). These defects suggest that early morphogenic movements in the lower genitourinary tract are disrupted in the absence of GATA5. GATA5 and GATA6 are coexpressed in the developing urogenital ridge but do not seem to have entirely overlapping functions during development of the female genitourinary system.[9]

GATA6 is expressed during both testicular and ovarian fetal development (Figure 3.2).[63] In the developing gonads, GATA4 and GATA6 have overlapping, but distinct expression patterns, which suggest different roles for these transcription factors. In addition, it is also possible that these factors complement each other's functions because GATA4 and GATA6 are expressed in similar cell types in the testis and ovary.[101,102]

Loss of GATA6 expression has been found in ovarian cancer and has been associated with hypoacetylation of histones H3 and H4 and loss of H3K4me3 at the promoter region.[90] Downregulation of GATA6 expression results in nuclear deformation and aneuploidy of ovarian surface epithelial cells.[103] In contrast to other cancers, these data indicate a tumour suppressor role for GATA6 in ovarian cancer. Tumor suppressing activities are also suggested for GATA4 and GATA5 whereas introduction of these genes into ovarian tumor cell lines greatly inhibits cell growth and survival.[104]

During pronephros formation human GATA3 expression is already detected in the nephric duct (Figure 3.2).[105] Subsequently, ureter tips and the collecting duct system of the metanephros are formed, which both show GATA3 expression.[106] Inactivation of the murine *GATA3* locus results in a morphologically abnormal nephric duct with an aberrant elongation path, loss of ureteric bud and a severe growth disturbance of de mesonephros due to the disturbance of a regulatory cascade consisting of GATA3 with β-catenin as upstream regulator and *Ret* as downstream target.[107]

In humans, *GATA3* haploinsufficiency leads to the HDR syndrome, a rare and complex disease characterised by the combination of HDR, associated with *GATA3* mutations (Table 3.3, Supplemental Table 3.1).[108] The majority of these mutations leads to loss of DNA binding caused by a disrupted ZnF2, or altered FOG2 interaction and/or DNA binding affinity by a disrupted ZnF1 (Table 3.3). Most of the HDR probands without *GATA3* mutations do not have renal abnormalities and no *GATA3* mutations are found

in patients with isolated hypoparathyreoidism.[109] This suggests that *GATA3* mutations are highly penetrant and result in the HDR phenotype. In addition, *GATA3*$^{+/-}$ mice show small size parathyroids resulting in failure to correct hypocalcaemia similar to HDR patients.[110] When *GATA3* is specifically deleted in the developing inner ear, defective formation of the cochlear prosensory domain and loss of spiral ganglion neurons is shown.[111] However, the exact mechanisms leading to the HDR phenotype remain to be elucidated.

Respiratory tract

The mammalian lung develops from budding of the foregut endoderm, in which both GATA4 and GATA6 are expressed. In vitro analysis of lung development of *GATA4*$^{ki/ki}$ mice show abnormal lobar development, revealing GATA4 as a candidate for FOG2-mediated early pulmonary development.[112] GATA6-regulated Wnt signalling controls the balance between bronchioalveolar stem cell expansion and epithelial differentiation required for both lung development and regeneration after lung injury.[113]

However, data about defects in GATA factors in lung diseases are scarce. Recently, GATA2 requirement for oncogenic Kras-driven lung tumorigenesis was reported. Moreover, inhibition of GATA2 regulated pathways in mice with KRAS mutant non-small cell lung cancer results in tumour regression.[114] Finally, a lung cancer susceptibility locus downstream of *GATA3* was identified.[115]

Mammary gland

Using *GATA3/LacZ* knock-in mice, GATA3 expression is observed at the earliest stages of embryonic mammary development (Figure 3.2). During puberty GATA3 is expressed in the terminal-end buds and within the adult mammary gland only in luminal epithelial cells. Targeted *GATA3* deletion at different stages of the embryonic mammary development showed loss or absence of mammary primordia and nipples.[116] Postnatal *GATA3* deletion resulted in loss of mammary gland development, and diminished expression of luminal differentiation markers, which indicates an important role of GATA3 in the luminal epithelium.[116,117] Loss of oestrogen receptor α (ERα) expression is observed in both *GATA3* knock-out mice and *FOG-2* knock-out mice.[117] Involvement of GATA3 and ERα in a positive cross-regulatory loop, which has been shown in breast cancer, may be an explanation for these phenomena.[118] Collectively, these data show that GATA3 is essential during embryonic development as well as the postnatal occurring morphogenesis.[116] Furthermore, GATA3 directs luminal differentiation of progenitor cells and is needed for active maintenance of the differentiated luminal phenotype.[117]

The crucial role of GATA3 in the mammary gland is further demonstrated by the observation of *GATA3* mutations in ~10% of human breast cancers. The spectrum of somatic mutations is diverse and cluster predominantly in the vicinity of the highly

conserved C-terminal second zinc-finger (Table 3.3; Supplemental Table 3.1).[119] Restoration of *GATA3* in breast cancer cell lines leads to differentiation, suppressed tumor dissemination[120], slower growth rates and induction of genes involved in luminal cell differentiation.[121] Thereby, GATA3 expression leads to reduced breast tumor outgrowth and inhibits pulmonary metastasis due to repression of metastasis-associated genes.[122] Recently it was described that GATA3 induces miR-29b expression, which in turn represses metastasis by changing tumor microenvironment.[123] Together these data indicate that *GATA3* might function as a tumor suppressor gene. In vitro- and in vivo data support this potential tumor suppressor function because loss of GATA3 leads to tumor progression and tumor dissemination in a murine luminal breast cancer model.[120] Prognostic and predictive features of GATA3 as a biomarker in breast cancer are discussed below in the clinical applications section.

Central Nervous System (CNS)

GATA2 is expressed early during CNS development in murine embryos (Figure 3.2).[124] Despite early lethality of *GATA2*$^{-/-}$ embryos (Table 3.2), several studies show that GATA2 is required for the development of sympathetic neurons[125], serotonergic hindbrain neurons[126], GABAergic midbrain neurons[127], retinorecipient neurons[128] and for the generation and cell fate determination of V2b spinal interneurons[129]. *GATA2*$^{-/-}$ embryos lack both GATA2 and GATA3 expression in the CNS, which indicates dependence of GATA3 expression on functional GATA2 during early differentiation of the neural tube.[130] The expression pattern of GATA3 during brain development is very similar to GATA2. *GATA3*$^{-/-}$ murine embryos also die early during embryonic development (Table 3.2) and have severe abnormalities of the brain and spinal cord.[7] Loss of *GATA3* results in reduced Th (tyrosine hydroxylase) and Dbh (dopamine β-hydroxylase) transcripts, which consequently leads to noradrenaline deficiency in the SNS. Administration of catecholamine intermediates to pregnant female *GATA3*$^{+/-}$ mice rescues *GATA3*$^{-/-}$ murine embryos, thereby partially unraveling the *GATA3* loss-induced lethality.[131] A transcriptional network, which includes GATA3[132], is essential for cell survival and differentiation of sympathetic neurons during embryonic development as well as during adult life.[133]

GATA4 is expressed in the embryonic and adult CNS and acts as a negative regulator of astrocyte proliferation and growth (Figure 3.2).[134] In the adult mouse and human, GATA6 is expressed in neurons, astrocytes, choroids plexus epithelium and endothelial cells (Figure 3.2).[135]

Loss of expression of GATA4 and GATA6 occurs in glioblastoma multiforme (GBM). Both *GATA4/6* gene promoters were found to be methylated and for *GATA4* also somatic mutations were found.[136,137] Limited evidence indicates that GATA4 regulates apoptosis-related genes in cultured GBM cell lines.[136] *GATA6* was identified in a mouse astrocytoma model as a novel tumor suppressor gene. Knockdown of GATA6 expression

in *RasV12* or *p53*$^{-/-}$ astrocytes led to acceleration of tumorigenesis. Mutations of *GATA6* occur during malignant progression of murine and human astrocytomas.[135]

Regulation of GATA genes and proteins in disease

Although mainly *GATA* gene mutations have been described above, chromosomal alterations as well as regulation of *GATA* genes and proteins on transcriptional and post-transcriptional levels can also contribute to disease development.

Recently it has been shown that combined *tet methylcytosine dioxygenase 2* (*TET2*) and *fms related tyrosine kinase 3* (*FLT3*) mutations regulate epigenetic silencing of *GATA2* by promotor hypermethylation in human AML.[138] In clear cell renal cell carcinomas downregulation of GATA3 expression by promoter hypermethylation results in decreased expression of TbetaRIII, a protein with tumor suppressor features, during disease progression.[139] Presence of suppressive histone (H3K27) trimethylation of *GATA3* together with absence of the GATA3 protein in anaplastic large cell lymphoma implicates epigenetical contribution in the pathogenesis of this disease.[140] Clues about the transcriptional regulation of the *GATA4* and *GATA6* genes come from a SUMO-specific protease 2 (SENP2) knockout model. These mice have reduced expression of GATA4 and GATA6 and defects in the embryonic heart. In SENP2 deficient embryos sumoylation of CBX4, accumulates and occupies the promoters of *GATA4* and *GATA6*, thereby leading to transcriptional repression.[141]

GATA4 is located at chromosome 8p, a chromosomal locus frequently deleted in multiple tumor types such as colorectal and oesophageal cancer.[142,143] Alternatively *GATA4* can be downregulated via epigenetic silencing, such as hypoacetylation of histones H3 and H4[90] and promoter CpG island hypermethylation, which has been observed in colorectal, gastric, oesophageal, lung, ovarian and HPV-driven oropharyngeal cancer, in GBM and in diffuse large B-cell lymphoma.[78,80,88,89,104,136,144,145] In contrast, *GATA4* amplification is recently described in certain gastric cancers which indicates a more oncogenic function.[92] Further studies are needed to unravel the molecular mechanisms of *GATA4* amplified in comparison with *GATA4* methylated gastric cancers.

GATA5 is located at chromosome 20q13, a locus which is often amplified and methylated in multiple cancer types. No coding sequence mutations in *GATA4* and *GATA5* have been described so far in colorectal- and breast cancer.[146,147] However, promoter methylation of *GATA5* might be established in order to downregulate increased gene expression imposed by amplification. Identified post-transcriptional modifications on GATA proteins include acetylation, phosphorylation and methylation (Figure 3.1). Protein stability of GATA2 and GATA3 is regulated by phosphorylation and ubiquitilation. Phosphorylation of GATA3 by respectively Cyclin-dependent kinase 1 (CDK1) and CDK2 was required for F-box/WD repeat-containing protein 7 (Fbw)-7 mediated ubiquitilation and degradation and contributed to precise differentiation of HSCs and T-cell lineages.[148,149] How GATA acetylation influences transcriptional

processes has been investigated for GATA1. It turns out that bromodomain protein Brd3 binds to acetylated GATA1 to regulate the chromatin occupancy at erythroid target genes.[150] For GATA4, post-transciptional modifications have mainly been studied in the context of hypertrophy of the heart. Activation of GATA4 occurs in part through acetylation by the transcriptional coactivator p300. Takaya et al. identified 4 GATA4 lysine residues that, when mutated, lacked p300-induced acetylation, DNA binding and transcriptional activities (Figure 3.1).[151] Phosphorylation of p300 by Cdk9 increases the ability of p300 to induce acetylation and DNA binding of GATA4.[152] Alternatively, phosphorylation of GATA4 on serine 105 is critical for a productive cardiac hypertrophic response to stress stimulation in adult mice.[153] Deacetylation of GATA4, and subsequent suppression of transcriptional activation, is mediated by histone deacetylase 2 (HDAC2) and the small homeodomain factor Hopx.[154] Recently it was reported that the GATA4 protein is methylated by Polycomb-repressive complex 2 member Ezh2. This reduced the interaction with and acetylation by p300, thereby reducing GATA4's transcriptional activity.[155] Together, this emphasises how important post-transciptional modifications are for the regulation of GATA activity.

Clinical applications of GATA transcription factor alterations

The above mentioned alterations in GATA factors might be applicable as biomarkers for early detection, diagnosis and prediction of prognosis and response to therapy.

Early detection markers. Non-invasive early diagnosis of CRC reduces mortality of this disease.[156] We have shown that *GATA4* promoter methylation is highly prevalent in CRC, suggesting that methylation is an early event in colorectal carcinogenesis. *GATA4* methylation, detected in fecal DNA has potential to be used as a biomarker for improving pre-selection tests for colonoscopy[80], especially if the clinical and analytical sensitivity and specificity can be improved by adding additional biomarkers and by introducing sensitive analysis techniques such as for example methylation on beads technology.[157]

Diagnostic markers. The expression of several GATA factors can be helpful in establishing a correct diagnosis. In ovarian cancer loss of GATA4 precedes loss of GATA6 expression and can differentiate between histological subtypes. Loss of both GATA4 and GATA6 expression is found in serous, clear cell and endometrioid ovarian cancer, but their expression can be detected in mucinous carcinomas.[158]

Prognostic markers. As already described above, *GATA1* mutations are found in nearly all AMKL patients with Down syndrome and are already detectable in the precursor lesion TMD. In addition, Down syndrome-neonates without *GATA1* mutations do not develop AMKL.[159,160] Together, the presence of *GATA1* mutations in Down syndrome-children might be a potential prognostic marker for identifying infants at higher risk of

developing AMKL.[161] Besides having a clinical value in AMKL, prognostic properties of GATA transcription factors are also described in T-ALL. Inherited genetic *GATA3* variants are identified in Philadelphia-like ALL (an ALL subtype with a poor prognosis) and are associated with early treatment response and a higher risk of relapse.[162]

GATA3 downregulation has been observed in ER-negative breast cancers and has been described as a strong prognostic indicator of breast cancer. Low GATA3 expression was strongly associated with aggressive disease and poor survival.[117] Vice versa, breast cancers expressing GATA3- and estrogen regulated genes exhibit a good prognosis and have better relapse-free and overall survival.[163] GATA3 has been considered to be a better prognostic marker for disease-free survival than commonly used variables such as ER status[164] although conflicting data have been published. However, GATA3 expression is highly correlated with the luminal A subtype which has a relatively favourable outcome compared with luminal B and basal-like subtypes.[165] An explanation could be the downregulation of $p18^{INK4C}$ transcription by GATA3 resulting in expansion of luminal progenitor cells thereby favouring the development of luminal type breast cancer.[166]

Recent studies indicate that *GATA2* may be a useful biomarker for predicting prognosis in AML. *GATA2* mutations are frequent in patients with a biallelic *CEBPA* mutation and are associated with a better survival.[167]

In oropharyngeal carcinomas, a methylation signature of 5 gene promoters, including *GATA4*, correlates with improved survival.[144] Eventually, loss of expression of GATA4 in GBM is associated with unfavourable patient survival.[136]

Recently it has been described that low GATA6 expression in lung adenocarcinomas is linked to increased incidence of metastasis and poor outcome.[168]

Predictive markers. Whole genome sequencing of samples from patients with ER-positive breast cancer, participating in aromatase inhibitor clinical trials identified 18 significantly mutated genes, including *GATA3*. Mutant *GATA3* correlated with suppression of proliferation upon aromatase inhibitor treatment and might therefore be a positive predictive marker for aromatase inhibitor response.[169]

Re-expression of GATA4 in GBM cells conferred sensitivity to temozolomide, a DNA alkylating agent used in GBM therapy.[136]

Recently, *GATA5* methylation was described as a potential predictive marker for patients with high-risk non-muscle-invasive bladder tumours. These patients had a better survival after treatment with Bacillus Calmette-Guérin (BCG) when *GATA5* was methylated.[170]

Therapeutic interventions. For regenerative medicine the generation of functional differentiated cell types is of great therapeutic interest. Since heart disease occurs frequently and the heart has little regenerative capacity after damage, procedures are sought that can transdifferentiate fibroblast into cardiac myocytes. A cocktail of

transcription factors, including GATA4 converts cardiac non-myocytes into cardiomyocyte-like cells in vivo, and alleviates cardiac injury.[171,172] Also in mouse liver engineering experiments GATA4 was one of the essential factors that contributed to the conversion of fibroblasts into functional hepatocyte-like cells.[173] These induced cells were able to restore liver function in half of fumarylacetoacetatehydrolase-deficient mice. *GATA4* is thus one of the pivotal genes that in combination with other transcription factors can be utilised to improve heart and liver function after damage. These promising results are the first steps for bringing regenerative medicine to the clinic. More knowledge of the different GATA protein functions and their downstream target genes is necessary before therapeutic strategies can be developed.

Conclusions and future perspectives

An increasing number of studies are being published, describing expression and function of *GATA* genes during development in different species.

Causal relationships between aberrations in *GATA* genes and several human diseases have become apparent. Numerous mutations in the *GATA* genes have been described above. Many disease-associated mutations are located in and around the Zinc finger regions. As those mutations are not specifically limited to the two Zinc fingers themselves, it is clear that the whole region is important for the proteins to be fully operational. Most likely mutations hinder the correct folding of the proteins and thereby obstruct GATA proteins from binding to their relevant binding partners. The application of next-generation sequencing technologies through whole-genome, whole-exome and whole-transcriptome approaches allows for substantial advances, which is expected to reveal more disease-associated alterations within *GATA* genes.

A better understanding of the regulation of GATA factors on transcriptional, translational and post-translational levels will give more leads to how GATAs can be used as biomarkers. Prospective clinical trials, based on these data, are necessary to determine the translational value of *GATA* genes as biomarkers.

References

1. Simon MC. Gotta have GATA. *Nat Genet* 1995;11:9-11.
2. He C, Cheng H, Zhou R. (2007) GATA family of transcription factors of vertebrates: phylogenetics and chromosomal synteny. *J Biosci* 2007;32:1273-1280.
3. Lowry JA, Atchley WR. Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. *J Mol Evol* 2000;50:103-115.
4. Morrisey EE, et al. GATA-4 activates transcription via two novel domains that are conserved within the GATA-4/5/6 subfamily. *J Biol Chem* 1997;272:8515-8524.
5. Fujiwara Y, et al. (1996) Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci U S A* 1996;93:12355-12358.
6. Tsai FY, et al. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* 1994;371:221-226.
7. Pandolfi PP, et al. Targeted disruption of the GATA3 gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis. *Nat Genet* 1995;11:40-44.
8. Kuo CT, et al. GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev* 1997;11:1048-1060.
9. Molkentin JD, et al. Abnormalities of the genitourinary tract in female mice lacking GATA5. *Mol Cell Biol* 2000;20:5256-5260.
10. Morrisey EE, et al. GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev* 1998;12:3579-3590.
11. Martin DI, et al. Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages. *Nature* 1990;344:444-447.
12. Simon M.C, et al. Rescue of erythroid development in gene targeted GATA-1- mouse embryonic stem cells. *Nat Genet* 1992;1:92-98.
13. Yu C, et al. Targeted deletion of a high-affinity GATAbinding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage in vivo. *J Exp Med* 2002;195:1387-1395.
14. Shimizu R, et al. In vivo requirements for GATA-1 functional domains during primitive and definitive erythropoiesis. *EMBO J* 2001;20:5250-5260.
15. Takai J, et al. The Gata1 5′ region harbors distinct cisregulatory modules that direct gene activation in erythroid cells and gene inactivation in HSCs. *Blood* 2013;122:3450-3460.
16. Weiss MJ, Keller G, Orkin SH. Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes Dev* 1994;8:1184-1197.
17. Grass J.A, et al. GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proc Natl Acad Sci U S A* 2003;100: 8811-8816.
18. Fujiwara Y, et al. Functional overlap of GATA-1 and GATA-2 in primitive hematopoietic development. *Blood* 2004;103:583-585.
19. Hollanda LM, et al. An inherited mutation leading to production of only the short isoform of GATA-1 is associated with impaired erythropoiesis. *Nat Genet* 2006;38:807-812.
20. Ludwig LS, et al. Altered translation of GATA1 in Diamond-Blackfan anemia. *Nat Med* 2014;20:748-753.
21. Shivdasani RA, et al. A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *EMBO J* 1997;16:3965-3973.
22. Huang Z, et al. (2007) STAT1 promotes megakaryopoiesis downstream of GATA-1 in mice. *J Clin Invest* 2007;117:3890-3899.
23. Muntean AG,. Crispino JD. Differential requirements for the activation domain and FOG-interaction surface of GATA-1 in megakaryocyte gene expression and development. *Blood* 2005;106:1223-1231.
24. Pimkin M, et al. Divergent functions of hematopoietic transcription factors in lineage priming and differentiation during erythro-megakaryopoiesis. Genome Res 2014;24:1932-1944.
25. Wechsler J, et al. Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nat Genet* 2002;32:148-152.
26. Calligaris R, et al. Alternative translation initiation site usage results in two functionally distinct forms of the GATA-1 transcription factor. *Proc Natl Acad Sci U S A* 1995;92:11598-11602.

27. Li Z, et al. Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1. *Nat Genet* 2005;37:613-619.
28. Groet J, et al. Acquired mutations in GATA1 in neonates with Down's syndrome with transient myeloid disorder. *Lancet* 2003;361:1617-1620.
29. Hitzler JK, et al. GATA1 mutations in transient leukemia and acute megakaryoblastic leukemia of Down syndrome. *Blood* 2003;101:4301-4304.
30. Yoshida K, et al. The landscape of somatic mutations in Down syndrome-related myeloid disorders. *Nat Genet* 2013;45:1293-1299.
31. Nikolaev SI, et al. Exome sequencing identifies putative drivers of progression of transient myeloproliferative disorder to AMKL in infants with Down syndrome. *Blood* 2013;122:554-561.
32. Cabelof DC, et al. Mutational spectrum at GATA1 provides insights into mutagenesis and leukemogenesis in Down syndrome. *Blood* 2009;114:2753-2763.
33. Stankiewicz MJ, Crispino JD. AKT collaborates with ERG and GATA1s to dysregulate megakaryopoiesis and promote AMKL. *Leukemia* 2013;27:1339-1347.
34. Roy A, et al. Perturbation of fetal liver hematopoietic stem and progenitor cell development by trisomy 21. *Proc Natl Acad Sci U S A* 2012;109:17579-17584.
35. Chou ST, et al. Trisomy 21-associated defects in human primitive hematopoiesis revealed through induced pluripotent stem cells. *Proc Natl Acad Sci U S A* 2012;109:17573-17578.
36. Maclean GA, et al. Altered hematopoiesis in trisomy 21 as revealed through in vitro differentiation of isogenic human pluripotent cells. *Proc Natl Acad Sci U S A* 2012;109:17567-17572.
37. Satge D. Are GATA1 mutations occurring at random in Down syndrome transient leukemia? *Med Hypotheses* 2014;83:154-159.
38. Woo AJ, et al. Developmental differences in IFN signaling affect GATA1s-induced megakaryocyte hyperproliferation. *J Clin Invest* 2013;123:13.
39. Nichols KE, et al. Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nat Genet* 2000;24:266-270.
40. Freson K, et al. Platelet characteristics in patients with X-linked macrothrombocytopenia because of a novel GATA1 mutation. *Blood* 2001;98:85-92.
41. Yu C, et al. X-linked thrombocytopenia with thalassemia from a mutation in the amino finger of GATA-1 affecting DNA binding rather than FOG-1 interaction. *Blood* 2002;100:2040-2045.
42. Rodrigues NP, et al. GATA-2 regulates granulocytemacrophage progenitor cell function. *Blood* 2008;112:4862-4873.
43. Ling KW, et al. GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells. *J Exp Med* 2004;200:871-882.
44. Bresnick EH, et al. Master regulatory GATA transcription factors: mechanistic principles and emerging links to hematologic malignancies. *Nucleic Acids Res* 2012;40:5819-5831.
45. Zhang SJ, et al. Gain-of-function mutation of GATA-2 in acute myeloid transformation of chronic myeloid leukemia. *Proc Natl Acad Sci U S A* 2008;105:2076-2081.
46. Hahn CN, et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet* 2011;43:1012-1017.
47. Hyde RK, Liu PP. GATA2 mutations lead to MDS and AML. *Nat Genet* 2011;43:926-927.
48. Leonard MW, Lim KC, Engel JD. Expression of the chicken GATA factor family during early erythroid development and differentiation. *Development* 1993;119:519-531.
49. Ho IC, et al. Human GATA-3: a lineage-restricted transcription factor that regulates the expression of the T cell receptor alpha gene. *EMBO J* 1991;10:1187-1192.
50. Frelin C, et al. GATA-3 regulates the self-renewal of long-term hematopoietic stem cells. *Nat Immunol* 2013;14:1037-1044.
51. Fitch SR, et al. Signaling from the sympathetic nervous system regulates hematopoietic stem cell emergence during embryogenesis. *Cell Stem Cell* 2012;11:554-566.
52. Hosoya T, Maillard I, Engel JD. From the cradle to the grave: activities of GATA-3 throughout T-cell development and differentiation. *Immunol Rev* 2010;238:110-125.
53. Ho IC, Tai TS, Pai SY. GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation. *Nat Rev Immunol* 2009;9:125-135.

54. Hu G, et al. Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nat Immunol* 2013;14:1190-1198.
55. Wang Y, et al. GATA-3 controls the maintenance and proliferation of T cells downstream of TCR and cytokine signaling. *Nat Immunol* 2013;14:714-722.
56. Sanda T, et al. Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell* 2012;22:209-221.
57. Zhang J, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* 2012;481:157-163.
58. Brewer A, Pizzey J. GATA factors in vertebrate heart development and disease. *Expert Rev Mol Med* 2006;8:1-20.
59. Crispino JD, et al. Proper coronary vascular development and heart morphogenesis depend on interaction of GATA-4 with FOG cofactors. *Genes Dev* 2001;15:839-844.
60. Heineke J, et al. Cardiomyocyte GATA4 functions as a stress-responsive regulator of angiogenesis in the murine heart. *J Clin Invest* 2007;117:3198-3210.
61. Yamak A, et al. Cyclin D2 is a GATA4 cofactor in cardiogenesis. *Proc Natl Acad Sci U S A* 2014;111: 1415-1420.
62. McCulley DJ, Black BL. Transcription factor pathways and congenital heart disease. *Curr Top Dev Biol* 2012;100:253-277.
63. Morrisey EE, et al. GATA-5: a transcriptional activator expressed in a novel temporally and spatially-restricted pattern during embryonic development. *Dev Biol* 1997;183:21-36.
64. Laforest B, Andelfinger G, Nemer M. Loss of Gata5 in mice leads to bicuspid aortic valve. *J Clin Invest* 2011;121:2876-2887.
65. Singh MK, et al. Gata4 and Gata5 cooperatively regulate cardiac myocyte proliferation in mice. *J Biol Chem* 2010;285:1765-1772.
66. Yang YQ, et al. Mutational spectrum of the GATA5 gene associated with familial atrial fibrillation. *Int J Cardiol* 2012;157:305-307.
67. Jiang JQ, et al. Prevalence and spectrum of GATA5 mutations associated with congenital heart disease. *Int J Cardiol* 2013;165:570-573.
68. Morrisey EE, et al. GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Dev Biol* 1996;177:309-322.
69. Lepore JJ, et al. GATA-6 regulates semaphorin 3C and is required in cardiac neural crest for cardiovascular morphogenesis. *J Clin Invest* 2006;116:929-939.
70. Tian Y, et al. Characterization and in vivo pharmacological rescue of a Wnt2-Gata6 pathway required for cardiac inflow tract development. *Dev Cell* 2010;18:275-287.
71. Kodo K, et al. GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling. *Proc Natl Acad Sci U S A* 2009;106:13933-13938.
72. Fujikura J, et al. Differentiation of embryonic stem cells is induced by GATA factors. *Genes Dev* 2002;16:784-789.
73. Soudais C, et al. Targeted mutagenesis of the transcription factor GATA-4 gene in mouse embryonic stem cells disrupts visceral endoderm differentiation in vitro. *Development* 1995;121:3877-3888.
74. Boudreau F, et al. Hepatocyte nuclear factor-1 alpha, GATA-4, and caudal related homeodomain protein Cdx2 interact functionally to modulate intestinal gene transcription. Implication for the developmental regulation of the sucraseisomaltase gene. *J Biol Chem* 2002;277:31909-31917.
75. Battle MA, et al. GATA4 is essential for jejunal function in mice. *Gastroenterology* 2008;135:1676-1686 e1.
76. Beuling E, et al. Conditional Gata4 deletion in mice induces bile acid absorption in the proximal small intestine. *Gut* 2010;59:888-895.
77. Beuling E, et al. GATA factors regulate proliferation, differentiation, and gene expression in small intestine of mature mice. *Gastroenterology* 2011;140:1219-1229 e1-2.
78. Akiyama Y, et al. GATA-4 and GATA-5 transcription factor genes and potential downstream antitumor target genes are epigenetically silenced in colorectal and gastric cancer. *Mol Cell Biol* 2003;23: 8429-8439.
79. Wen XZ, et al. Methylation of GATA-4 and GATA-5 and development of sporadic gastric carcinomas. *World J Gastroenterol* 2010;16:1201-1208.

80. Hellebrekers DM, et al. GATA4 and GATA5 are potential tumor suppressors and biomarkers in colorectal cancer. *Clin Cancer Res* 2009;15:3990-3997.
81. Watt AJ, et al. Development of the mammalian liver and ventral pancreas is dependent on GATA4. *BMC Dev Biol* 2007;7:37.
82. Zheng R, et al. Function of GATA factors in the adult mouse liver. *PloS One* 2013;8:e83723.
83. Zhao R, et al. GATA6 is essential for embryonic development of the liver but dispensable for early heart formation. *Mol Cell Biol* 2005;25:2622-2631.
84. Parviz F, et al. Hepatocyte nuclear factor 4alpha controls the development of a hepatic epithelium and liver morphogenesis. *Nat Genet* 2003;34:292-296.
85. Lango Allen H, et al. GATA6 haploinsufficiency causes pancreatic agenesis in humans. *Nat Genet* 2012;44:20-22.
86. Martinelli P, et al. Gata6 is required for complete acinar differentiation and maintenance of the exocrine pancreas in adult mice. *Gut* 2013;62:1481-1488.
87. Kwei KA, et al. Genomic profiling identifies GATA6 as a candidate oncogene amplified in pancreatobiliary cancer. *PLoS Genet* 2008;4:e1000081.
88. Guo M, et al. Hypermethylation of the GATA genes in lung cancer. Clin Cancer Res 2004;10:7917-7924.
89. Guo M, et al. Hypermethylation of the GATA gene family in esophageal cancer. *Int J Cancer* 2006;119:2078-2083.
90. Caslini C, et al. Histone modifications silence the GATA transcription factor genes in ovarian cancer. *Oncogene* 2006;25:5446-5461.
91. Gao X, et al. Distinct functions are implicated for the GATA-4, -5, and -6 transcription factors in the regulation of intestine epithelial cell differentiation. *Mol Cell Biol* 1998;18:2901-2911.
92. Chia NY, et al. Regulatory crosstalk between lineagesurvival oncogenes KLF5, GATA4 and GATA6 cooperatively promotes gastric cancer development. *Gut* 2014;64:707-719.
93. Ito E, et al. Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis. *Nature* 1993;362:466-468.
94. Lindeboom F, et al. A tissue-specific knockout reveals that Gata1 is not essential for Sertoli cell function in the mouse. *Nucleic Acids Res* 2003;31:5405-5412.
95. Viger RS, et al. Transcription factor GATA-4 is expressed in a sexually dimorphic pattern during mouse gonadal development and is a potent activator of the Mullerian inhibiting substance promoter. *Development* 1998;125:2665-2675.
96. Hu YC., Okumura LM, Page DC. Gata4 is required for formation of the genital ridge in mice. *PLoS Genet* 2013;9:e1003629.
97. Tevosian SG, et al. Gonadal differentiation, sex determination and normal Sry expression in mice require direct interaction between transcription partners GATA4 and FOG2. *Development* 2002;129: 4627-4634.
98. Bouma GJ, et al. Correct dosage of Fog2 and Gata4 transcription factors is critical for fetal testis development in mice. *Proc Natl Acad Sci U S A* 2007;104:14994-14999.
99. Gierl MS, et al. GADD45 G functions in male sex determination by promoting p38 signaling and Sry expression. *Dev Cell* 2012;23:1032-1042.
100. Lourenco D, et al. Loss-of-function mutation in GATA4 causes anomalies of human testicular development. *Proc Natl Acad Sci U S A* 2011;108:1597-1602.
101. Ketola I, et al. Expression and regulation of transcription factors GATA-4 and GATA-6 in developing mouse testis. *Endocrinology* 1999;140:1470-1480.
102. Heikinheimo M, et al. Expression and hormonal regulation of transcription factors GATA-4 and GATA-6 in the mouse ovary. *Endocrinology* 1997;138:3505-3514.
103. Capo-chichi CD, et al. Loss of GATA6 leads to nuclear deformation and aneuploidy in ovarian cancer. *Mol Cell Biol* 2009;29:4766-4777.
104. Wakana K, et al. Involvement of GATA-4/-5 transcription factors in ovarian carcinogenesis. *Cancer Letters* 2006;241:281-288.
105. George KM, et al. Embryonic expression and cloning of the murine GATA-3 gene. *Development* 1994;120:2673-2686.
106. Grote D, et al. Pax 2/8-regulated Gata 3 expression is necessary for morphogenesis and guidance of the nephric duct in the developing kidney. *Development* 2006;133:53-61.

107. Chia I, et al. Nephric duct insertion is a crucial step in urinary tract maturation that is regulated by a Gata3-Raldh2-Ret molecular network in mice. *Development* 2011;138:2089-2097.
108. Van Esch H, et al. GATA3 haplo-insufficiency causes human HDR syndrome. Nature 2000;406:419-422.
109. Ali A, et al. Functional characterization of GATA3 mutations causing the hypoparathyroidism-deafness-renal (HDR) dysplasia syndrome: insight into mechanisms of DNA binding by the GATA3 transcription factor. *Hum Mol Genet* 2007;16:265-275.
110. Grigorieva IV, et al. Gata3-deficient mice develop parathyroid abnormalities due to dysregulation of the parathyroid-specific transcription factor Gcm2. *J Clin Invest* 2010;120:2144-2155.
111. Luo XJ, et al. GATA3 controls the specification of prosensory domain and neuronal survival in the mouse cochlea. *Hum Mol Genet* 2013;22:3609-3623.
112. Ackerman KG, et al. Gata4 is necessary for normal pulmonary lobar development. *Am J Respir Cell Mol Biol* 2007;36:391-397.
113. Zhang Y, et al. A Gata6-Wnt pathway required for epithelial stem cell development and airway regeneration. *Nat Genet* 2008;40:862-870.
114. Kumar MS, et al. The GATA2 transcriptional network is requisite for RAS oncogene-driven non-small cell lung cancer. *Cell* 2012;149:642-655.
115. Dong J, et al. Association analyses identify multiple new lung cancer susceptibility loci and their interactions with smoking in the Chinese population. *Nat Genet* 2012;44:895-899.
116. Asselin-Labat ML, et al. Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol* 2007;9:201-209.
117. Kouros-Mehr H, et al. GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* 2006;127:1041-1055.
118. Eeckhoute J, et al. Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer. *Cancer Res* 2007;67:6477-6483.
119. Koboldt DC, et al. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;490: 61-70.
120. Kouros-Mehr H, et al. GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model. *Cancer Ce*ll 2008;13:141-152.
121. Usary J, et al. Mutation of GATA3 in human breast tumors. *Oncogene* 2004;23:7669-7678.
122. Dydensborg AB, et al. GATA3 inhibits breast cancer growth and pulmonary breast cancer metastasis. *Oncogene* 2009;28:2634-2642.
123. Chou J, et al. GATA3 suppresses metastasis and modulates the tumour microenvironment by regulating microRNA-29b expression. *Nat Cell Biol* 2013;15:201-213.
124. ZhouW, et al. Counting alleles to predict recurrence of early-stage colorectal cancers. *Lancet* 2002;359: 219-225.
125. Tsarovina K, et al. Essential role of Gata transcription factors in sympathetic neuron development. *Development* 2004;131:4775-4786.
126. Craven SE, et al. Gata2 specifies serotonergic neurons downstream of sonic hedgehog. *Development* 2004;131:1165-1173.
127. Kala K, et al. Gata2 is a tissue-specific postmitotic selector gene for midbrain GABAergic neurons. *Development* 2009;136:253-262.
128. Willett RT, Greene LA. Gata2 is required for migration and differentiation of retinorecipient neurons in the superior colliculus. *J Neurosci* 2011;31:4444-4455.
129. Zhou Y, Yamamoto M, Engel JD. GATA2 is required for the generation of V2 interneurons. *Development* 2000;127:3829-3838.
130. Karunaratne A, et al. GATA proteins identify a novel ventral interneuron subclass in the developing chick spinal cord. *Dev Biol* 2002;249:30-43.
131. Lim KC, et al. Gata3 loss leads to embryonic lethality due to noradrenaline deficiency of the sympathetic nervous system. *Nat Genet* 2000;25:209-212.
132. Goridis C, Rohrer H. Specification of catecholaminergic and serotonergic neurons. *Nat Rev Neurosci* 2002;3:531-541.
133. Tsarovina K, et al. The Gata3 transcription factor is required for the survival of embryonic and adult sympathetic neurons. *J Neurosci* 2010;30:10833-10843.

134. Agnihotri S, et al. GATA4 is a regulator of astrocyte cell proliferation and apoptosis in the human and murine central nervous system. *Oncogene* 2009;28:3033-3046.
135. Kamnasaran D, et al. GATA6 is an astrocytoma tumor suppressor gene identified by gene trapping of mouse glioma model. *Proc Natl Acad Sci U S A* 2007;104:8053-8058.
136. Agnihotri S, et al. A GATA4-regulated tumor suppressor network represses formation of malignant human astrocytomas. *J Exp Med* 2011;208:689-702.
137. Martinez R, et al. A microarray-based DNA methylation study of glioblastoma multiforme. *Epigenetics* 2009;4:255-264.
138. Shih AH, et al. Mutational cooperativity linked to combinatorial epigenetic gain of function in acute myeloid leukemia. *Cancer Cell* 2015;27:502-515.
139. Cooper SJ, et al. Loss of type III transforming growth factor-beta receptor expression is due to methylation silencing of the transcription factor GATA3 in renal cell carcinoma. *Oncogene* 2010;29: 2905-2915.
140. Joosten M, et al. Histone acetylation and DNA demethylation of T cells result in an anaplastic large cell lymphoma-like phenotype. *Haematologica* 2013;98:247-254.
141. Kang X, et al. SUMO-specific protease 2 is essential for suppression of polycomb group protein-mediated gene silencing during embryonic development. *Mol Cell* 2010;38:191-201.
142. Derks S, et al. Promoter methylation precedes chromosomal alterations in colorectal cancer development. *Cell Oncol* 2006;28:247-257.
143. Lin L, et al. A minimal critical region of the 8p22-23 amplicon in esophageal adenocarcinomas defined using sequence tagged site-amplification mapping and quantitative polymerase chain reaction includes the GATA-4 gene. *Cancer Res* 2000;60:1341-1347.
144. Kostareli E, et al. HPV-related methylation signature predicts survival in oropharyngeal squamous cell carcinomas. *J Clin Invest* 2013;123:2488-2501.
145. Pike B.L, et al. DNA methylation profiles in diffuse large B-cell lymphoma and their relationship to gene expression status. *Leukemia* 2008;22:1035-1043.
146. Sjoblom T, et al. The consensus coding sequences of human breast and colorectal cancers. *Science* 2006;314:268-274.
147. Wood LD, et al. The genomic landscapes of human breast and colorectal cancers. *Science* 2007;318: 1108-1113.
148. Kitagawa K, et al. Fbw7 targets GATA3 through cyclin-dependent kinase 2-dependent proteolysis and contributes to regulation of T-cell development. *Mol Cell Biol* 2014;34:2732-2744.
149. Nakajima T, et al. Regulation of GATA-binding protein 2 levels via ubiquitin-dependent degradation by Fbw7: involvement of cyclin B-cyclin-dependent kinase 1-mediated phosphorylation of THR176 in GATA-binding protein 2. *J Biol Chem* 2015;290:10368-10381.
150. Lamonica JM, et al. Bromodomain protein Brd3 associates with acetylated GATA1 to promote its chromatin occupancy at erythroid target genes. *Proc Natl Acad Sci U S A* 2011;108:E159-E168.
151. Takaya T, et al. Identification of p300-targeted acetylated residues in GATA4 during hypertrophic responses in cardiac myocytes. *J Biol Chem* 2008;283:9828-9835.
152. Sunagawa Y, et al. Cyclin-dependent kinase-9 is a component of the p300/GATA4 complex required for phenylephrine-induced hypertrophy in cardiomyocytes. *J Biol Chem* 2010;285:9556-9568.
153. van Berlo JH, et al. Serine 105 phosphorylation of transcription factor GATA4 is necessary for stress-induced cardiac hypertrophy in vivo. *Proc Natl Acad Sci U S A* 2011;108:12331-12336.
154. Trivedi CM, et al. Hopx and Hdac2 interact to modulate Gata4 acetylation and embryonic cardiac myocyte proliferation. *Dev Cell* 2010;19:450-459.
155. He A, et al. PRC2 directly methylates GATA4 and represses its transcriptional activity. *Genes Dev* 2012;26:37-42.
156. Hewitson P, et al. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. *Am J Gastroenterol* 2008;103:1541-1549.
157. Guzzetta AA, et al. The promise of methylation on beads for cancer detection and treatment. *Exp Rev Mol Diagn* 2014;14:845-852.
158. Cai KQ, et al. Loss of GATA4 and GATA6 expression specifies ovarian cancer histological subtypes and precedes neoplastic transformation of ovarian surface epithelia. *PloS One* 2009;4:e6454.

159. Pine SR, et al. Incidence and clinical implications of GATA1 mutations in newborns with Down syndrome. *Blood* 2007;110:2128-2131.
160. Roberts I, et al. GATA1-mutant clones are frequent and often unsuspected in babies with Down syndrome: identification of a population at risk of leukemia. *Blood* 2013;122:3908-3917.
161. Roy A, et al. Acute megakaryoblastic leukaemia (AMKL) and transient myeloproliferative disorder (TMD) in Down syndrome: a multi-step model of myeloid leukaemogenesis. *Br J Haematol* 2009;147:3-12.
162. Perez-Andreu V, et al. Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. *Nat Genet* 2013;45:1494-1498.
163. Oh DS, et al. Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. *J Clin Oncol* 2006;24:1656-1664.
164. Mehra R, et al. Identification of GATA3 as a breast cancer prognostic marker by global gene expression metaanalysis. *Cancer Res* 2005;65:11259-11264.
165. Albergaria A, et al. Expression of FOXA1 and GATA-3 in breast cancer: the prognostic significance in hormone receptor-negative tumours. *Breast Cancer Res* 2009;11:R40.
166. Pei XH, et al. CDK inhibitor p18(INK4c) is a downstream target of GATA3 and restrains mammary luminal progenitor cell proliferation and tumorigenesis. *Cancer Cell* 2009;15:389-401.
167. Fasan A, et al. GATA2 mutations are frequent in intermediate-risk karyotype AML with biallelic CEBPA mutations and are associated with favorable prognosis. *Leukemia* 2013;27:482-485.
168. Cheung WK, et al. Control of alveolar differentiation by the lineage transcription factors GATA6 and HOPX inhibits lung adenocarcinoma metastasis. *Cancer Cell* 2013;23:725-738.
169. Ellis MJ, et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* 2012;486:353-360.
170. Agundez M, et al. Evaluation of the methylation status of tumour suppressor genes for predicting bacillus Calmette-Guerin response in patients with T1G3 high-risk bladder tumours. *Eur Urol* 2011;60:131-140.
171. Song K, et al. Heart repair by reprogramming nonmyocytes with cardiac transcription factors. *Nature* 2012;485:599-604.
172. Qian L, et al. In vivo reprogramming of murine cardiac fibroblasts into induced cardiomyocytes. *Nature* 2012;485:593-598.
173. Huang P, et al. Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* 2011;475:386-389.

Supplementary table

Table S3.1 Disease related mutations of GATA1 – GATA6 zinc fingers with clinical phenotype from previously reported studies as depicted in Figure 3.1.

Gene	Amino acid change	Phenotype	Literature reference
GATA1	Val205Met	XLT	1
GATA1	Gly208Arg	XLT, XLT with MDS	2,3,4
GATA1	Gly208Ser	XLT	5
GATA1	Arg216Gln	XLTT	6,7,8,9
GATA1	Arg216Trp	CEP	10
GATA1	Asp218Gly	XLT	11
GATA1	Asp218Tyr	XLT	12
GATA2	Arg293Gln	AML	13
GATA2	Arg307Leu	AML	14
GATA2	Arg307Trp	AML	15
GATA2	Arg308Pro	AML	14,15,16
GATA2	Asn317His	AML	13
GATA2	Asn317Ile	AML	14,15
GATA2	Asn317Ser	AML	14
GATA2	Ala318Gly	AML	13
GATA2	Ala318fs	MDS with myeloid transformation	17
GATA2	Ala318Thr	AML	13,15
GATA2	Ala318Val	(paediatric) AML	13-15,18
GATA2	Gly320Asp	AML	13,14
GATA2	Gly320Val	AML	15
GATA2	Leu321Arg	AML	14
GATA2	Leu321His	AML	14
GATA2	Leu321Phe	AML	13,-15
GATA2	Leu321Pro	AML	13,15
GATA2	Leu321Val	AML	13,14
GATA2	Gln328Pro	AML	13
GATA2	Asn329Gln	AML	13- 15
GATA2	Arg330Leu	AML	15
GATA2	Arg330Pro	AML	15
GATA2	Arg330*	MDS with myeloid transformation AML	17,19
GATA2	Leu332fs	Emberger syndrome	20
GATA2	Arg337*	Emberger syndrome	21
GATA2	Arg337*	MonoMAC MDS Emberger syndrome	17,21,22
GATA2	Ser340_Asn381del	MDS MonoMAC	23
GATA2	Ala341fs	Emberger syndrome	21,23
GATA2	Ala342fs	NK-cell deficiency	24
GATA2	Ala342Thr	Paediatric AML	25
GATA2	Ala350_Asn351ins8	AML	16
GATA2	Thr354Lys	AML	14
GATA2	Thr354Met	MonoMAC AML MDS NK-cell deficiency	17,20,22-24,26-31

Table S3.1 (continued)

Gene	Amino acid change	Phenotype	Literature reference
GATA2	Thr355del	Familial MDS	27
GATA2	Thr357Ser	Paediatric AML	25
GATA2	Thr358Lys	AML/MDS	32
GATA2	Leu359Val	AML/MDS	32-34
		acute myeloid transformation of CML	
		CML	
GATA2	Arg361Cys	Emberger syndrome	17,22,23
		MonoMAC and MDS	
		MDS with myeloid transformation	
GATA2	Arg361His	AML	14,17
		MDS with myeloid transformation	
GATA2	Arg361Leu	Emberger syndrome	21
GATA2	Arg362_Asn365del	MDS with myeloid transformation	24,29,35
		DCML	
GATA2	Arg362Gln	(paediatric) AML	14,18,25
GATA2	Arg362Gly	Paediatric AML	18,25
GATA2	Arg362Pro	Paediatric AML	18,25
GATA2	Arg362fs	Paediatric AML	18
GATA2	Asp367fs	Monocytopenia	17,22
		MDS with myeloid transformation	
		MonoMAC	
GATA2	Asn371Lys	MDS with myeloid transformation	17,29
		DCML	
GATA2	Ala372Thr	AML	19
GATA2	Cys373Arg	Emberger syndrome	21
GATA2	Cys373_Tyr377del	MDS with myeloid transformation	17
GATA2	Leu375Ile	Paediatric AML	25
GATA2	Leu379Gln	AML	14
GATA2	Met388Thr	MonoMAC	17,22
		Monocytopenia	
		Laryngeal cancer	
GATA2	Met388Val	DCML	19
GATA2	Lys390del	MDS	23
GATA2	Arg396Gln	MonoMAC	17,19,22,31,36
		DCML and MDS	
		MDS with myeloid transformation	
		AML	
		Low B-cell	
GATA2	Arg396Glu	MDS	26
		AML	
GATA2	Arg396Trp	MonoMAC and MDS	17
		MDS with myeloid transformation	
GATA2	Arg398Gln	MonoMAC	23
GATA2	Arg398Trp	DCML	17,22-24,31,37
		MonoMAC	
		MDS (with myeloid transformation)	
		CMML	
		NK-cell deficiency	
GATA3	Arg261Gly	HDR syndrome	38
GATA3	Ser270fs	T-ALL	39
GATA3	Thr271Ile	HDR syndrome	40,41

Table S3.1 (continued)

Gene	Amino acid change	Phenotype	Literature reference
GATA3	Trp274Arg	HDR syndrome	42
GATA3	Trp274Leu	HDR syndrome	43
GATA3	Arg275Gln	T-ALL	39
GATA3	Arg275Trp	T-ALL	39
GATA3	Arg276Pro	HDR syndrome	44
GATA3	Arg276*	HDR syndrome	45,46
GATA3	Asn285Thr	T-ALL	39
GATA3	Met293Lys	Breast cancer	47-49
GATA3	Lys302*	HDR syndrome	38,50
GATA3	Ser308fs	Breast cancer	51
GATA3	Ser308*	Breast cancer	51
GATA3	Ala309_Ala313del	T-ALL	39
GATA3	Thr315_Ala318del	HDR syndrome	45
GATA3	Thr315fs	Breast cancer	51
GATA3	Cys317fs	HDR syndrome	52
GATA3	Cys317Arg	HDR syndrome	53
GATA3	Cys317Ser	HDR syndrome	38
GATA3	Asn319Lys	HDR syndrome	53
GATA3	Cys320Ser	HDR syndrome	54
GATA3	Thr322fs	Breast cancer	49
GATA3	Arg329fs	Breast cancer	47,48,55
GATA3	Arg329del	Breast cancer	55
GATA3	Arg330Lys	Breast cancer	47,48
GATA3	Asn331fs	Breast cancer	47,48
GATA3	Ala332fs	Breast cancer	47,48,56
GATA3	Asp335fs	Breast cancer	56
GATA3	Cys341Tyr	HDR syndrome	57
GATA3	Leu343Phe	Breast cancer	51
GATA3	Tyr345fs	Breast cancer	49
GATA3	Leu347Val	T-ALL	39
GATA3	Leu347Arg	HDR syndrome	46
GATA3	Arg352Ser	HDR syndrome	58
GATA3	Leu354fs	Breast cancer	49
GATA3	Leu354*	HDR syndrome	38,46
GATA3	Met356*	HDR syndrome	42,46
GATA3	Met356fs	Breast cancer	49
GATA3	Lys357*	Breast cancer	49
GATA3	Glu359fs	Breast cancer	49
GATA3	Arg364Gly	Breast cancer	49
GATA3	Arg364Ser	Breast cancer	47,48
GATA3	Arg366Leu	Breast cancer	51
GATA3	Arg366*	HDR syndrome Breast cancer	42,48,51,53,59
GATA3	Ser369*	HDR syndrome	53
GATA4	Glu216Asp	CHD (TOF)	60
GATA4	Gly221Arg	Anomalies of human testicular development	61
GATA4	Met223Thr	CHD (VSD)	62
GATA4	Pro226=	CHD (VSD)	63
GATA4	Pro226fs	CHD (AVSD)	62
GATA4	Pro226Gln	CHD (DCM)	64
GATA4	Arg229Ser	CHD (VSD / ASD / AVSD)	62

Table S3.1 (continued)

Gene	Amino acid change	Phenotype	Literature reference
GATA4	Thr233=	CHD (DORV / PFO / TOF / PA / VSD / AS)	65-68
GATA4	Gly234Ser	CHD (AVSD)	62
GATA4	Asn239=	CHD (VSD)	62
GATA4	Asn239Ser	CHD (VSD)	62
GATA4	Cys241=	CHD (TGA / VSD / ASD)	63,66
GATA4	Tyr244Cys	CHD (VSD / AVSD)	62
GATA4	Tyr244=	CHD (AVSD / TOF /PA / LSVC)	68
GATA4	Met247Thr	CHD (AF)	69
GATA4	Asn248=	CHD (ASD / VSD)	63
GATA4	Asn248Ser	CHD (ASD / AVSD)	62
GATA4	Ile250Asn	CHD (VSD)	63
GATA4	Arg252Pro	CHD (AVSD)	62
GATA4	Ile255Thr	CHD (ASD)	62
GATA4	Arg260Gln	CHD (VSD)	62
GATA4	Leu261Pro	CHD (VSD / ASD)	62
GATA4	Ser262=	CHD (TOF / PA)	68
GATA4	Ala263Gly	CHD (VSD)	70
GATA4	Arg266*	CHD (AVSD)	62
GATA4	Val267Met	CHD (ASD / PDA)	71,72
GATA4	Cys271Ser	CHD (DCM)	73
GATA4	Asn273Lys	Pancreatic agenesis	74
GATA4	Asn273Ser	CHD (AVSD)	62
GATA4	Cys274=	CHD (TOF / PA / DORV / VSD / TGA / ASD)	66,67
GATA4	Thr277Ile	CHD (AVSD)	62
GATA4	Thr279Ser	CHD (DCM)	64
GATA4	Thr280Met	CHD (ASD)	75
GATA4	Arg283His	CHD (AVSD)	62
GATA4	Asn285Ser	CHD (TOF)	76
GATA4	Asn285Lys	CHD (AVSD)	62
GATA4	Val291Leu	CHD (DCM)	77
GATA4	Cys292Arg	CHD (VSD / ASD / AVSD)	62,78
GATA4	Ala294Val	CHD (ASD)	62
GATA4	Gly296Arg	CHD (VSD)	79
GATA4	Gly296Cys	CHD (ASD / PS)	80
GATA4	Gly296Ser	CHD (ASD / VSD / AVSD / PS)	81,82
GATA4	His302Arg	CHD (AVSD)	62
GATA4	Met310Val	CHD (ASD)	83
GATA4	Gln316Glu	CHD (ASD)	67
GATA4	Arg318Trp	CHD (ASD) Pancreatic agenesis	84
GATA4	Lys319Glu	CHD (ASD / PS)	85
GATA4	Leu325=	CHD (ASD)	63
GATA4	Lys329Asn	CHD	86
GATA5	Arg187Gly	CHD (TOF)	87
GATA5	Val190Ala	CHD (VSD / TOF)	88
GATA5	Leu199Val	CHD (VSD)	89
GATA5	Trp200Gly	CHD (AF)	90
GATA5	Asp203Glu	CHD (TOF)	91
GATA5	Asp203=	CHD (BAV)	92
GATA5	His207Arg	CHD (TOF)	87
GATA5	Tyr208*	CHD (TOF)	91

Table S3.1 (continued)

Gene	Amino acid change	Phenotype	Literature reference
GATA5	Cys210Gly	CHD (AF)	93
GATA5	Lys218Thr	CHD (AF)	94
GATA5	Leu226=	CHD (BAV)	92
GATA5	Leu233Pro	CHD (BAV)	95,96
GATA5	Thr252Pro	CHD (BAV)	97
GATA5	Ala266Pro	CHD (AF / TOF)	88,94
GATA5	His274Arg	CHD (VSD)	88
GATA5	Lys284=	CHD (BAV)	92
GATA6	Gly394Cys	CHD (TOF)	98
GATA6	Asp404Tyr	CHD (TOF)	99
GATA6	Cys447Arg	Diabetes	100
GATA6	Thr452Ala	Pancreatic agenesis CHD (ASD)	101
GATA6	Arg456Cys	Pancreatic agenesis CHD (VSD / PTA / TOF) ConHD	101,102
GATA6	Arg456His	Pancreatic agenesis CHD (e.g. PTA and VSD)	101
GATA6	Glu460*	CHD (TOF)	99
GATA6	Asn466Asp	Pancreatic agenesis PTA	101
GATA6	Asn466His	CHD (PTA)	103
GATA6	Asn466Ser	CHD Permanent neonatal diabetes	104
GATA6	Ala467Thr	Pancreatic agenesis CHD (ASD / PS)	101
GATA6	Gly469Glu	Pancreatic agenesis	100
GATA6	Gly469Val	CHD (AF)	105
GATA6	Lys473Gln	Pancreatic agenesis CHD (ASD)	101
GATA6	Arg479Gly	Pancreatic agenesis CHD (e.g. AVSD, PS)	100
GATA6	Met483fs	Pancreatic agenesis CHD	101
GATA6	Arg493*	Pancreatic agenesis CHD	106
GATA6	Lys500fs	Pancreatic agenesis CHD (e.g. VSD and PS)	101
GATA6	Lys502fs	Pancreatic agenesis CHD (VSD / PTA)	107

AF atrial fibrillation; *AML* acute myeloid leukemia; *AS* aortic stenosis; *ASD* atrial septal defect; *AVSD* atrial ventricular septal defect; *BAV* bicuspid aortic valve; *CEP* congenital erythropoietic porphyria; *CHD* congenital heart disease; *CML* chronic myeloid leukemia; *CMML* chronic myelomonocytic leukemia; *ConHD* congenital hernia diaphragmatica; *DCM* dilated cardiomyopathy; *DCML* dendritic cell, monocyte, B lymphocyte, and natural killer lymphocyte deficiency; *DORV* double-outlet right ventricle; *HDR* syndrome hypoparathyroidism, sensorineural deafness and renal insufficiency; *MDS* myelodysplastic syndrome; *PA* pulmonary atresia; *PDA* patent ductus arteriosis; *PFO* patent foramen ovale; *PS* pulmonary valve stenosis; *T-ALL* T-cell acute lymphoblastic leukemia; *TGA* transposition of the great arteries; *TOF* Tetralogy of Fallot; *VSD* ventricular septal defect; *XLTT* X-linked thrombocytopenia with thalassemia; *XLT* X-linked thrombocytopenia.

References

1. Nichols KE, et al. Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nat Genet* 2000;24(3):266-270.
2. Duhrsen U, et al. Long-term outcome of hemizygous and heterozygous carriers of a germline GATA1 (G208R) mutation. *Ann Hemat* 2011;90(3):301-306
3. Kratz CP, et al. Congenital transfusion-dependent anemia and thrombocytopenia with myelodysplasia due to a recurrent GATA1(G208R) germline mutation. *Leukemia* 22(2), 432-434
.4. Del Vecchio GC, et al. Dyserythropoietic anemia and thrombocytopenia due to a novel mutation in GATA-1. *Acta Haematol* 2005;114(2):113-136.
5. Mehaffey MG, et al. X-linked thrombocytopenia caused by a novel mutation of GATA-1. *Blood* 2001;98(9):2681-2688.
6. Yu C, et al. X-linked thrombocytopenia with thalassemia from a mutation in the amino finger of GATA-1 affecting DNA binding rather than FOG-1 interaction. *Blood* 2002;00(6):2040-2045.
7. Tubman VN, et al. X-linked gray platelet syndrome due to a GATA1 Arg216Gln mutation. *Blood* 2007;109(8):3297-3299.
8. Balduini CL, et al. Effects of the R216Q mutation of GATA-1 on erythropoiesis and megakaryocytopoiesis. *Thromb Haemost* 2004;91(1):129-140.
9. Hughan SC, et al. Selective impairment of platelet activation to collagen in the absence of GATA1. *Blood* 2005;105(11):4369-4376.
10. Phillips JD, et al. Congenital erythropoietic porphyria due to a mutation in GATA1: the first trans-acting mutation causative for a human porphyria. *Blood* 2007;109(6):2618-2621.
11. Freson K, et al. Platelet characteristics in patients with X-linked macrothrombocytopenia because of a novel GATA1 mutation. *Blood* 2001;98(1):85-92.
12. Freson K, et al. Different substitutions at residue D218 of the X-linked transcription factor GATA1 lead to altered clinical severity of macrothrombocytopenia and anemia and are associated with variable skewed X inactivation. *Hum Mol Genet* 2002;11(2):147-152.
13. Greif PA, et al. GATA2 zinc finger 1 mutations associated with biallelic CEBPA mutations define a unique genetic entity of acute myeloid leukemia. *Blood* 2012;120(2):395-403.
14. Fasan A, et al. GATA2 mutations are frequent in intermediate-risk karyotype AML with biallelic CEBPA mutations and are associated with favorable prognosis. *Leukemia* 2013;27(2):482-485.
15. Green CL, et al. GATA2 mutations in sporadic and familial acute myeloid leukaemia patients with CEBPA mutations. *Br J Haematol* 2013;161(5):701-705.
16. Niimi K, et al. GATA2 zinc finger 2 mutation found in acute myeloid leukemia impairs myeloid differentiation. *Leuk Res Rep* 2013;2(1):21-25.
17. West RR, et al. Acquired ASXL1 mutations are common in patients with inherited GATA2 mutations and correlate with myeloid transformation. *Haematologica* 2014;99(2):276-281.
18. Luesink M, et al. High GATA2 expression is a poor prognostic marker in pediatric acute myeloid leukemia. *Blood* 2012;120(10):2064-2075.
19. Pasquet M, et al. High frequency of GATA2 mutations in patients with mild chronic neutropenia evolving to MonoMac syndrome, myelodysplasia, and acute myeloid leukemia. *Blood* 2013;121(5): 822-829.
20. Kazenwadel J, et al. Loss-of-function germline GATA2 mutations in patients with MDS/AML or MonoMAC syndrome and primary lymphedema reveal a key role for GATA2 in the lymphatic vasculature. *Blood* 2012;119(5):1283-1291.
21. Ostergaard P, et al. Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat Genet* 2011;43(10):929-931.
22. Hsu AP, et al. GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. *Blood* 121(19), 3830-7, S1-7
23. Dickinson RE, et al. The evolution of cellular deficiency in GATA2 mutation. *Blood* 2013;123(6):863-874.
24. Mace EM, et al. Mutations in GATA2 cause human NK cell deficiency with specific loss of the CD56(bright) subset. *Blood* 2013;121(14):2669-2677.

25. Shiba N, et al. Mutations of the GATA2 and CEBPA genes in paediatric acute myeloid leukaemia. *Br J Haematol* 2014;164(1):142-145.
26. Holme H, et al. Marked genetic heterogeneity in familial myelodysplasia/acute myeloid leukaemia. *Br J Haematol* 2012;158(2):242-248.
27. Hahn CN, et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet* 2011;43(10):1012-1017.
28. Bodor C, et al. Germ-line GATA2 p.THR354MET mutation in familial myelodysplastic syndrome with acquired monosomy 7 and ASXL1 mutation demonstrating rapid onset and poor survival. *Haematologica* 2012;97(6):890-894.
29. Hsu AP, et al. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. Blood 2011;118(10):2653-2655.
30. Bigley V, et al. The human syndrome of dendritic cell, monocyte, B and NK lymphoid deficiency. *J Exp Med* 2011;208(2):227-234.
31. Cuellar-Rodriguez J, et al. Successful allogeneic hematopoietic stem cell transplantation for GATA2 deficiency. *Blood* 2011;118(13):3715-3720.
32. Gao J, et al. Heritable GATA2 mutations associated with familial AML-MDS: a case report and review of literature. *J Hematol Oncol* 2014;7(1):36.
33. Zhang SJ, et al. Gain-of-function mutation of GATA-2 in acute myeloid transformation of chronic myeloid leukemia. *Proc Natl Acad Sci U S A* 2008;105(6):2076-2081.
34. Zhang SJ, Shi JY, Li JY. GATA-2 L359 V mutation is exclusively associated with CML progression but not other hematological malignancies and GATA-2 P250A is a novel single nucleotide polymorphism. *Leuk Res* 2009;33(8):1141-1143.
35. West ES, et al. Generalized verrucosis in a patient with GATA2 deficiency. *Br J Dermatol* 2014;170(5):1182-1186.
36. Ishida H, et al. GATA-2 anomaly and clinical phenotype of a sporadic case of lymphedema, dendritic cell, monocyte, B- and NK-cell (DCML) deficiency, and myelodysplasia. *Eur J Pediatr* 2012;171(8): 1273-1276.
37. Dickinson RE, et al. Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood* 2011;118(10):2656-2658.
38. Nakamura A, et al. Molecular analysis of the GATA3 gene in five Japanese patients with HDR syndrome. *Endocr J* 2011;58(2):123-130.
39. Zhang J, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* 2012;481(7380):157-163.
40. Gomes TS, et al. HDR syndrome: a follow-up genotype-phenotype analysis of a de novo missense Thr272Ile mutation in exon 4 of GATA3. *Klinische Padiatrie* 2012;224(7):452-454.
41. Gaynor KU, et al. A missense GATA3 mutation, Thr272Ile, causes the hypoparathyroidism, deafness, and renal dysplasia syndrome. *J Clin Endocrinol Metab* 2009;94(10):3897-3904.
42. Muroya K, et al. GATA3 abnormalities and the phenotypic spectrum of HDR syndrome. *J Med Genet* 2001;38(6):374-380.
43. Fukami M, et al. GATA3 abnormalities in six patients with HDR syndrome. *Endocr J* 2011;58(2):117-121.
44. Zahirieh A, et al. Functional analysis of a novel GATA3 mutation in a family with the hypoparathyroidism, deafness, and renal dysplasia syndrome. *J Clin Endocrinol Metab* 2005;90(4): 2445-2450.
45. Van Esch H, et al. GATA3 haplo-insufficiency causes human HDR syndrome. *Nature* 2000;406(6794):419-422.
46. Ali A, et al. Functional characterization of GATA3 mutations causing the hypoparathyroidism-deafness-renal (HDR) dysplasia syndrome: insight into mechanisms of DNA binding by the GATA3 transcription factor. *Hum Mol Genet* 2007;16(3):265-275.
47. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;490(7418):61-70.
48. Jiang YZ, et al. GATA3 mutations define a unique subtype of luminal-like breast cancer with improved survival. *Cancer* 2014;120(9):1329-1337.
49. Ellis MJ, et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* 2012;486(7403):353-360.

50. Mino Y, et al. Identification of a novel insertion mutation in GATA3 with HDR syndrome. *Clin Exp Nephrol* 2005;9(1):58-61.
51. Usary J, et al. Mutation of GATA3 in human breast tumors. Oncogene 2004;23(46):7669-7678.
52. Chenouard A, et al. Renal phenotypic variability in HDR syndrome: glomerular nephropathy as a novel finding. *Eur J Pediatr* 2013;172(1):107-110
53. Nesbit MA, et al. Characterization of GATA3 mutations in the hypoparathyroidism, deafness, and renal dysplasia (HDR) syndrome. *J Biol Chem* 279(21), 22624-34
54. Ohta M, et al. Novel dominant-negative mutant of GATA3 in HDR syndrome. *J Mol Med* 2011;89(1): 43-50.
55. Gaynor KU, et al. GATA3 mutations found in breast cancers may be associated with aberrant nuclear localization, reduced transactivation and cell invasiveness. *Horm Cancer* 2013;4(3):123-139.
56. Wheler JJ, et al. Unique molecular signatures as a hallmark of patients with metastatic breast cancer: implications for current treatment paradigms. *Oncotarget* 2014;5(9):2349-2354.
57. Moldovan O, et al. A new case of HDR syndrome with severe female genital tract malformation: comment on "Novel mutation in the gene encoding the GATA3 transcription factor in a Spanish familial case of hypoparathyroidism, deafness, and renal dysplasia (HDR) syndrome with female genital tract malformations" by Hernandez et al. *Am J Med Genet* 2011;155A(9):2329-2330.
58. Chiu WY, et al. Identification of three novel mutations in the GATA3 gene responsible for familial hypoparathyroidism and deafness in the Chinese population. *J Clin Endocrinol Metab* 2006;91(11): 4587-4592.
59. Sun Y, et al. Germinal mosaicism of GATA3 in a family with HDR syndrome. Am J Med Genet 2009; 149A(4), 776-8
60. Nemer G, et al. A novel mutation in the GATA4 gene in patients with Tetralogy of Fallot. *Hum Mutat* 2006;27(3):293-294.
61. Lourenco D, et al. Loss-of-function mutation in GATA4 causes anomalies of human testicular development. *Proc Natl Acad Sci U S A* 2011;108(4):1597-1602.
62. Reamon-Buettner SM, Borlak J. GATA4 zinc finger mutations as a molecular rationale for septation defects of the human heart. *J Med Genet* 2005;42(5):e32
63. Wang E, et al. Identification of functional mutations in GATA4 in patients with congenital heart disease. *PloS One* 2013;8(4):e62138.
64. Li J, et al. Prevalence and spectrum of GATA4 mutations associated with sporadic dilated cardiomyopathy. *Gene* 2014;548(2):174-81.
65. Reamon-Buettner SM, Cho SH, Borlak J. Mutations in the 3'-untranslated region of GATA4 as molecular hotspots for congenital heart disease (CHD). *BMC Med Genet* 2007;8:38.
66. Butler TL, et al. GATA4 mutations in 357 unrelated patients with congenital heart malformation. *Genet Test Mol Biomarkers* 2010;14(6):797-802.
67. Tomita-Mitchell A, et al. GATA4 sequence variants in patients with congenital heart disease. *J Med Genet* 2007;44(12):779-783.
68. Schluterman MK, et al. Screening and biochemical analysis of GATA4 sequence variations identified in patients with congenital heart disease. *Am J Med Genet* 2007;143A(8):817-823.
69. Posch MG, et al. Mutations in the cardiac transcription factor GATA4 in patients with lone atrial fibrillation. *Eur J Med Genet* 2010;53(4):201-203.
70. Xiong F, et al. Analyses of GATA4, NKX2.5, and TFAP2B genes in subjects from southern China with sporadic congenital heart disease. *Cardiovasc Pathol* 2013;22(2):141-145.
71. Wang J, et al. [Genetic screening for novel GATA4 mutations associated with congenital atrial septal defect]. *Zhonghua Xin Xue Guan Bing Za Zhi* 2010;38(5):429-434.
72. Tang ZH, et al. Two novel missense mutations of GATA4 gene in Chinese patients with sporadic congenital heart defects. *Zhonghua Xin Xue Guan Bing Za Zhi* 2006;23(2):134-137.
73. Li RG, et al. GATA4 loss-of-function mutation underlies familial dilated cardiomyopathy. *Biochem Biophys Res Commun* 2013;439(4):591-596.
74. Shaw-Smith C, et al. GATA4 mutations are a cause of neonatal and childhood-onset diabetes. *Diabetes* 2014;63(8):2888-2894.
75. Chen Y, et al. A novel mutation of GATA4 in a familial atrial septal defect. *Clin Chim Acta* 2010;411(21-22):1741-1745.

76. Yang YQ, et al. GATA4 loss-of-function mutations underlie familial tetralogy of fallot. *Hum Mutat* 2013;34(12):1662-1671.
77. Zhao L, et al. A novel GATA4 loss-of-function mutation responsible for familial dilated cardiomyopathy. *Int J Mol Med* 2014;33(3):654-660.
78. Reamon-Buettner SM, Spanel-Borowski K, Borlak J. Bridging the gap between anatomy and molecular genetics for an improved understanding of congenital heart disease. *Ann Anat* 2006;188(3):213-220.
79. Wang J, et al. A novel GATA4 mutation responsible for congenital ventricular septal defects. *Int J Mol Med* 2011;28(4):557-564.
80. Rajagopal SK, et al. Spectrum of heart disease associated with murine and human GATA4 mutation. *J Mol Cell Cardiol* 2007;43(6):677-685.
81. Garg V, et al. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* 2003;424(6947):443-447.
82. Sarkozy A, et al. Spectrum of atrial septal defects associated with mutations of NKX2.5 and GATA4 transcription factors. *J Med Genet* 2005;42(2):e16.
83. Chen Y, et al. A novel mutation in GATA4 gene associated with dominant inherited familial atrial septal defect. *J Thorac Cardiovasc Surg* 2010;140(3):684-697.
84. D'Amato E, et al. Genetic investigation in an Italian child with an unusual association of atrial septal defect, attributable to a new familial GATA4 gene mutation, and neonatal diabetes due to pancreatic agenesis. *Diabet Med* 2010;27(10):1195-1200.
85. Xiang R, et al. A novel mutation of GATA4 (K319E) is responsible for familial atrial septal defect and pulmonary valve stenosis. *Gene* 2013;pii:S0378-1119(13)01423-6.
86. Wang J, et al. [Novel GATA4 mutations identified in patients with congenital heart disease]. *Zhonghua Yi Xue Za Zhi* 2010;90(10):667-671.
87. Wei D, et al. GATA5 loss-of-function mutations underlie tetralogy of fallot. *Int J Med Sci* 2013;10(1): 34-42.
88. Jiang JQ, et al. Prevalence and spectrum of GATA5 mutations associated with congenital heart disease. *Int J Cardiol* 2013;165(3):570-573.
89. Wei D, et al. GATA5 loss-of-function mutation responsible for the congenital ventriculoseptal defect. *Pediatr Cardiol* 2013;34(3);504-511.
90. Wang XH, et al. A novel GATA5 loss-of-function mutation underlies lone atrial fibrillation. *Int J Mol Med* 2013;31(1):43-50.
91. Huang RT, et al. Somatic GATA5 mutations in sporadic tetralogy of Fallot. *Int J Mol Med* 2014;33(5):1227-1235.
92. Foffa I, et al. Sequencing of NOTCH1, GATA5, TGFBR1 and TGFBR2 genes in familial cases of bicuspid aortic valve. *BMC Med Genet* 2013;14:44.
93. Gu JY, et al. Novel GATA5 loss-of-function mutations underlie familial atrial fibrillation. Clinics 2012;67(12):1393-1399.
94. Yang YQ, et al. Mutational spectrum of the GATA5 gene associated with familial atrial fibrillation. *Int J Cardiol* 2012;157(2):305-307.
95. Francis C, et al. 95 Identification Of Likely Pathogenic Variants In Patients With Bicuspid Aortic Valve: Correlation Of Complex Genotype With A More Severe Aortic Phenotype. *Heart* 2014;100(Suppl 3): A55-A56.
96. Bonachea EM, et al. Rare GATA5 sequence variants identified in individuals with bicuspid aortic valve. *Pediatr Res* 2014;76(2):211-216.
97. Shi LM, et al. GATA5 loss-of-function mutations associated with congenital bicuspid aortic valve. *Int J Mol Med* 2014;33(5):1219-1226.
98. Huang RT, et al. Somatic mutations in the GATA6 gene underlie sporadic tetralogy of Fallot. *Int J Mol Med* 2013;31(1):51-58.
99. Wang J, et al. Novel GATA6 mutations associated with congenital ventricular septal defect or tetralogy of fallot. *DNA Cell Biol* 2012;31(11):1610-1617.
100. De Franco E, et al. GATA6 mutations cause a broad phenotypic spectrum of diabetes from pancreatic agenesis to adult-onset diabetes without exocrine insufficiency. *Diabetes* 2013;62(3):993-997.
101. Allen HL, et al. GATA6 haploinsufficiency causes pancreatic agenesis in humans. *Nat Genet* 2012;44(1):20-22.

102. Yu L, et al. Whole exome sequencing identifies de novo mutations in GATA6 associated with congenital diaphragmatic hernia. *J Med Genet* 2014;51(3):197-202.
103. Kodo K, et al. GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling. *Proc Natl Acad Sci U S A* 106(33), 13933-13938.
104. Catli G, et al. A novel GATA6 mutation leading to congenital heart defects and permanent neonatal diabetes: a case report. *Diabetes Metab* 2013;39(4):370-374.
105. Li J, et al. Novel GATA6 loss-of-function mutation responsible for familial atrial fibrillation. *Int J Mol Med* 2012;30(4):783-790.
106. Suzuki S, et al. A case of pancreatic agenesis and congenital heart defects with a novel GATA6 nonsense mutation: evidence of haploinsufficiency due to nonsense-mediated mRNA decay. *Am J Med Genet* 2014;164A(2):476-479.
107. Bonnefond A, et al. GATA6 inactivating mutations are associated with heart defects and, inconsistently, with pancreatic agenesis and diabetes. *Diabetologia* 2012;55(10):2845-2847.

Chapter 4

N-Myc downstream-regulated gene 4 (*NDRG4*): a candidate tumor suppressor gene and potential biomarker for colorectal cancer

Veerle Melotte, Marjolein HFM Lentjes, Sandra M van den Bosch, Debby MEI Hellebrekers, Joep PJ de Hoon, Kim AD Wouters, Kathleen LJ Daenen, Iris EJM Partouns-Hendriks, Filip Stessels, Joost Louwagie, Kim M Smits, Matty P Weijenberg, Silvia Sanduleanu, Carolina AJ Khalid-de Bakker, Frank A Oort, Gerrit A Meijer, Daisy MAE Jonkers, James G Herman, Adriaan P de Bruïne, Manon van Engeland

J Natl Cancer Inst 2009;101(13):916-927

Abstract

Background

Identification of hypermethylated tumor suppressor genes in body fluids is an appealing strategy for the noninvasive detection of colorectal cancer. Here we examined the role of N-Myc downstream-regulated gene 4 (*NDRG4*) as novel tumor suppressor and biomarker in colorectal cancer.

Methods

NDRG4 promoter methylation was analyzed in human colorectal cancer cell lines ,in colorectal cancer tissue and in non-cancerous colon mucosa by using methylation-specific polymerase chain reaction (PCR) and bisulfite sequencing. NDRG4 mRNA and protein expression were studied using reverse transcription–PCR and immunohistochemistry, respectively. Tumor suppressor functions of NDRG4 were examined by colony formation, cell proliferation, and migration and invasion assays in colorectal cancer cell lines that were stably transfected with a NDRG4 expression construct. Quantitative methylation-specific PCR was used to examine the utility of *NDRG4* promoter methylation as biomarker in fecal DNA from 75 colorectal cancer patients and 75 control subjects. All *P* values are two-sided.

Results

The prevalence of *NDRG4* promoter methylation in two independent series of colorectal cancers was 86% (71/83) and 70% (128/184) compared with 4% (2/48) in noncancerous colon mucosa ($p<.001$). NDRG4 mRNA and protein expression was decreased in colorectal cancer tissue compared with noncancerous colon mucosa. NDRG4 overexpression in colorectal cancer cell lines suppressed colony formation ($p=.014$) and cell proliferation ($p<.001$) and invasion ($p<.001$). *NDRG4* promoter methylation analysis in fecal DNA from a training set of colorectal cancer patients and control subjects yielded a sensitivity of 61% (95% CI = 43% to 79%) and a specificity of 93% (95% CI = 90% to 97%). An independent test set of colorectal cancer patients and control subjects yielded a sensitivity of 53% (95% CI = 39% to 67%) and a specificity of 100% (95% CI = 100% to 100%).

Conclusions

NDRG4 is a candidate tumor suppressor gene in colorectal cancer whose expression is frequently inactivated by promoter methylation. *NDRG4* promoter methylation is a potential biomarker for the noninvasive detection of colorectal cancer in stool samples.

Context and caveats

Prior knowledge

Identification of tumor suppressor gene promoter hypermethylation in fecal DNA is a promising strategy for noninvasive detection of colorectal cancer. N-Myc downstream-regulated gene 4 (*NDRG4*) is a potential tumor suppressor in colorectal cancer.

Study design

NDRG4 promoter methylation and expression were analyzed in human colorectal cancer cell lines, noncancerous colon mucosa, and colorectal cancer tissue. NDRG4 tumor suppressor functions were examined in colorectal cancer cells. *NDRG4* promoter methylation was examined as a potential biomarker in stool from colorectal cancer patients and subjects without colorectal cancer.

Contribution

NDRG4 promoter methylation was prevalent in colorectal cancers compared with noncancerous colon mucosa. NDRG4 mRNA and protein expression were decreased in colorectal cancer tissue compared with noncancerous colon mucosa. *NDRG4* overexpression in human colorectal cancer cells inhibited colony formation and cell proliferation and invasion in vitro. A methylation-specific polymerase chain reaction assay for *NDRG4* promoter methylation identified colorectal cancer when it was present (sensitivity) in 53% of colorectal cancer cases and correctly categorized a subject as cancer free (specificity) 100% of the time.

Implications

NDRG4 is a candidate tumor suppressor gene in colorectal cancer. *NDRG4* promoter methylation is a potential biomarker for the noninvasive detection of colorectal cancer in stool samples.

Limitations

Not all stool samples from colorectal cancer patients were collected before colonoscopy as was done for the control subjects. The colorectal cancer patients were older than the subjects without colorectal cancer

Introduction

Hypermethylation of CpG islands in the promoter region of genes is associated with gene silencing, may serve as a mechanism to inactivate tumor suppressor genes in colorectal cancer carcinogenesis, and can be analyzed easily by using methylation-specific polymerase chain reaction (PCR). Identification of methylation markers that are sensitive and specific for colorectal cancer detection may improve the early detection of this disease. Previous microarray experiments[1] to identify genes that are epigenetically regulated in tumor endothelium revealed 81 genes whose expression was decreased in tumor endothelial cells compared with quiescent endothelial cells and that were re-expressed after treatment with the DNA methylation inhibitor 5-aza-2'-deoxycytidine (DAC) and the histone deacetylase inhibitor trichostatin A (TSA). Silencing of these genes in tumor endothelium was associated with modifications of histones in the promoter regions but not with promoter CpG island methylation.[1] It is interesting that 21 (26%) of the 81 genes were reported to be hypermethylated at their promoters and silenced in various tumor types, suggesting that expression of those genes in tumor cells may be regulated by promoter methylation.[1]

One of the identified genes was N-myc downstream-regulated gene 4 (*NDRG4*; also known as *SMAP-8* and *BDM1*). The protein encoded by this gene, NDRG4, is a member of the NDRG protein family, which comprises four members named NDRG1–4 that have 57%–65% amino acid sequence homology.[2,3] NDRG1 is the most extensively studied member of the NDRG family. Expression of NDRG1 is decreased in cancer cells[4-9] but increases in cancer cells that are treated with DAC.[5,6] It has been demonstrated that NDRG1 overexpression in colorectal cancer cell lines reduces their ability to metastasize in nude mice, and the NDGR1-mediated suppression of metastasis is thought to involve the induction of colorectal cancer cell differentiation and a partial reversal of the metastatic phenotype.[6] *NDRG2* has been described as a candidate tumor suppressor gene[10,11] and displays promoter CpG island methylation in meningiomas[10] and in breast, liver, and lung cancer cell lines.[12] To our knowledge, the roles of NDRG3 and NDRG4 in cancer have not been addressed.

The *NDRG4* gene is located at chromosome 16q21–q22.3, spans 26 kilobases, and contains 17 exons that include the entire sequence of the three cDNA isoforms: *NDRG4-B*, *NDRG4-B*^{var}, and *NDRG4-H*.[2] To our knowledge, NDRG4 expression has only been described in brain and heart using northern blot analysis. The molecular characterization of NDRG4 and the role of this protein in the nervous system has been investigated mainly in the rat[13-16], where it is thought to participate in processes that lead to cellular differentiation and neurite formation.[14]

Here we examined the expression of NDRG4 at the mRNA and protein levels in normal human colon mucosa and human colorectal cancer tissue. In addition, we examined the mechanism underlying the decreased expression of *NDRG4* in colorectal cancer and investigated a possible tumor suppressor function of NDRG4 in colorectal

cancer cell lines by measuring colony formation, and cell proliferation, migration, and invasion in colorectal cancer cell lines that stably overexpressed *NDGR4*. Finally, we investigated the potential utility of *NDRG4* promoter methylation as a biomarker for early detection of colorectal cancer in stool.

Materials and methods

Study population and tissue samples

NDRG4 promoter methylation was investigated in two independent well-characterized tissue series from colorectal carcinoma patients, adenoma patients, and control subjects without cancer. The first hospital-based series consisted of formalin-fixed, paraffin-embedded colorectal cancer tissues (n=90) from patients who were older than 50 years at colorectal cancer diagnosis during 1995–2003 and were retrospectively retrieved from the tissue archive of the Department of Pathology at the Maastricht University Medical Center. We also retrieved noncancerous healthy colon mucosa (n=79) and adenoma (n=62) tissue from these patients when available. As control tissue, we used histologically normal biopsy material from control subjects who underwent endoscopy during 1987–2004 for nonspecific abdominal complaints (n=51), adenoma biopsy samples from patients diagnosed during 1988–1995 and who did not develop colorectal cancer within 10 years of the adenoma diagnosis (n=22), and resected colon mucosa from patients diagnosed during 1985–2004 with various inflammatory bowel conditions (n=33). The inflammatory bowel conditions in the latter group of control tissues included Crohn disease (n=1), colitis ulcerosa (n=5), nonspecific inflammation (n=9), and diverticulitis (n=18). Control tissues were excluded if the patient had been diagnosed with colorectal cancer in the past or during follow up. We excluded patients with and without colorectal cancer who had been diagnosed with additional cancers (excluding non-melanoma skin cancer). Characteristics of the study populations are shown in Supplementary Figure S4.1, Supplementary Table S4.1, and Supplementary Table S4.2. Numbers of samples in the results section may differ from those described in the table because not all samples could be amplified by using methylation-specific PCR.

The second population-based series of formalin-fixed, paraffin-embedded colorectal cancers (n = 184) was randomly selected from the prospective Netherlands Cohort Study on Diet and Cancer (NLCS), which has been described in detail elsewhere.[17,18] The 184 patients from whom this series of colorectal cancers were obtained were similar to the complete group of eligible colorectal cancer patients in the NLCS with respect to age at diagnosis, sex, TNM (tumor–node–metastasis) stage[19], and tumor location. This study was approved by the Medical Ethical Committee (MEC) of the Maastricht University Medical Center.

DNA Isolation

A 5-μm section of each tissue block was stained with hematoxylin and eosin and reviewed by the study pathologist (APdB). Five sections (20 μm thick) were deparaffinized and subjected to genomic DNA extraction by using a Puregene DNA isolation kit (Qiagen) according to the manufacturer's instructions.

Collection and Preparation of Fecal DNA

Stool samples were obtained from healthy colonoscopy-negative control subjects older than 50 years of age who underwent colonoscopy screening for colorectal cancer within the framework of a workplace-based community colorectal cancer study at the Maastricht University Medical Center. Stool samples and colorectal cancer tissues were collected from colonoscopy-confirmed colorectal cancer case patients who were diagnosed with all stages of colorectal cancer at the VU University Medical Center in Amsterdam. Two independent sets of noncancerous control subjects and colorectal cancer patients were collected; a training set consisting of 28 colorectal cancer patients and 45 noncancerous control subjects and a test set consisting of 47 colorectal cancer patients and 30 noncancerous control subjects. Series characteristics are shown in Supplementary Table S4.3. The MEC of the Maastricht University Medical Center and the Dutch Health Council approved this study. Written informed consent was obtained from all subjects who provided stool samples. All control stool samples from both sets of patients, one colorectal cancer stool samples of the training set and three colorectal cancer stool samples of the test set were collected within 2 weeks before colonic purgation and colonoscopy. Twenty-seven colorectal cancer stool samples from the training and forty-four colorectal cancer stool samples from the validation set were collected 5–7 days after colonoscopy and before resection of the tumor.

Stool stabilization buffer was added to the stool sample by the subject immediately after defecation (EXACT Sciences, Marlborough, MA) and stool samples were processed within 48 hours after defecation. For recovery of human DNA, whole stool samples were homogenized in a 7-fold excess volume of stool stabilization buffer and aliquoted in 32-ml portions that contained the equivalent of 4 g of stool each. Single aliquots were centrifuged at 13100*g* for 2 minutes, and the supernatants were incubated with RNase A (80 U/ml) for 60 minutes at 37°C. Total DNA was precipitated by using 2.2 ml of 3M sodium acetate (pH 5.2) and 22 ml of 100% isopropanol, centrifuged at 4500*g* for 5 minutes, and resuspended in 4 ml of 1 mM EDTA, 0.01M Tris-HCl (pH 7.4).

Half of each DNA sample was stored at −20°C and the other half was purified as follows. Stool lysis buffer (1.5 ml; ASL buffer, Qiagen) and an InhibitEX tablet (Qiagen) were added to 2 ml of the DNA sample and the mixture was centrifuged at 4500*g* for 5 minutes. After centrifugation, the supernatant was pipetted into a new tube and the pellet was discarded. We added 2 ml of the supernatant to 150 μl of proteinase K

(>600 mAU/ml, Qiagen), then added 2.4 ml of lysis buffer (AL buffer, Qiagen), and incubated the mixture for 10 minutes at 70°C. We next added 2 ml of ethanol to the incubated sample and the mixture was loaded onto a QIAamp Midi column (Qiagen), which was centrifuged at 1850*g* for 3 minutes. The column was washed sequentially with 2 ml of wash buffer 1 (AW1 buffer, Qiagen) and 2 ml of wash buffer 2 (AW2 buffer, Qiagen), with centrifugation at 4500*g* for 15 minutes. We added 200 μl of elution buffer (buffer AE; Qiagen) onto the membrane of the column, and the column was incubated at room temperature for 5 minutes. Finally, the column was centrifuged at 4500*g* for 2 minutes. The eluted fecal DNA (2 μg) was subjected to bisulfite modification in 96-well plates (Tecan) by using an EZ-96 DNA Methylation kit (Zymo Research) according to the manufacturer's protocol. Bisulfite-treated fecal DNA was concentrated by using a DNA Clean & Concentrator kit (Zymo Research Co). *NDRG4* promoter methylation in fecal DNA was analyzed by quantitative methylation-specific polymerase chain reaction (PCR) as described below.

Sodium bisulfite conversion, sequencing, and quantitative methylation-specific PCR

Sodium bisulfite modification, which converts unmethylated cytosine residues to uracil residues, was carried out on 500 ng genomic DNA isolated from the tissue sections and colorectal cancer cell lines (HT29, SW48, CaCo2, Colo205, RKO, LS174T, HCT116, and SW480) with the use of an EZ DNA methylation kit (ZYMO Research Co, Orange, CA) according to the manufacturer's instructions. *NDRG4* methylation-specific PCR analysis was performed on bisulfite-modified DNA as described in detail elsewhere.[20,21]

For sequencing, bisulfite-modified DNA was amplified using methylation-specific primers (shown in Supplementary Table S4.3) and a PCR profile consisting of an initial denaturation at 95°C for 5 minutes; followed by 35 cycles of 30 seconds at 95°C, 30 seconds at 60°C, and 1 minute at 72°C; and a final extension at 72°C for 5 minutes. PCR products were cloned by using a TOPO-TA cloning kit (Invitrogen, Breda, the Netherlands), and six independent bacterial clones were sequenced by using an automated DNA sequencer (Applied Biosystems, Foster City, CA). Quantitative methylation-specific PCR was performed by using a 7900HT real-time PCR system (Applied Biosystems) as follows: 2.4 μ bisulfite-modified DNA was added to a PCR mix containing buffer [16.6 mM $(NH_4)_2SO_4$, 67 mM Tris, 6,7 mM $MgCl_2$, 10 mM ß-mercaptoethanol], dATP, dCTP, dGTP, and dTTP (each at 5 mM), forward primer (6 ng/μl), reverse primer (18 ng/μl), a single-stranded oligonucleotide hybridization probe (0.16 μM), bovine serum albumin (BSA; 0.1 μg), and Jumpstart Taq polymerase (0.4 U; Sigma-Aldrich). Beta-actin was used as a reference gene for normalization. The PCR program was as follows: 5 minutes at 95°C; followed by 45 cycles of 30 seconds at 95°C, 30 seconds at 57°C, and 30 seconds at 72°C; followed by 5 minutes at 72°C. Serially diluted plasmids (20-2×10^6 copies) containing the target sequence were amplified to generate a standard curve against which the unknown samples are quantified by interpolation of their PCR cycle number (Ct value) to the corresponding

plasmid copy. Primer sequences are provided in Supplemental Table S4.3. Samples were handled and analyzed in a blinded fashion during storage, DNA isolation, and PCR analysis. One quantitative methylation-specific PCR experiment was performed for each independent set of patients.

Cell culture and transfections

Human colorectal cancer cell lines (HT29, SW48, CaCo2, Colo205, RKO, LS174T, HCT116, and SW480; all from LGC, Teddington, UK) were cultured in Dulbecco's modified Eagle medium (DMEM) (Invitrogen, Breda, the Netherlands) supplemented with 10% heat-inactivated fetal calf serum (FBS; HyClone, Etten-Leur, the Netherlands). To investigate the effect of re-expression of *NDRG4*, RKO and HCT116 cells were treated for 3 days with 1 μM 5-aza-2' deoxycytidine (DAC; Sigma). The full-length *NDRG4* complementary DNA (Origene) was subcloned into a pCMV6-Neo vector (Origene) to create pCMV6-NDRG4. HCT116 cells were transfected with pCMV6-NDRG4 or empty vector (pCMV6) by use of a Nucleofector Kit V (Amaxa Biosystems, Gaithersburg, MD) according to the manufacturer's guidelines. RKO cells were transfected with pCMV6-NDRG4 or pCMV6 by using Lipofectamine 2000 Reagent (Invitrogen) according to the manufacturer's protocol. Transfected HCT116 and RKO cells were grown for 10 days in medium containing G418 (at 400 μg/ml for HCT116 and 1 mg/ml for RKO; Invitrogen) to select for cells that were stably transfected with the pCMV6-based plasmids.

Quantitative real-time PCR

Total RNA was isolated from colorectal cancer cell lines and tissues from patients and control subjects by using a RNeasy Mini kit (Qiagen) according to the manufacturer's instructions and treated with RNase-free DNAse (Qiagen) to remove contaminating genomic DNA. cDNA was synthesized from 1 μg of the DNAse-treated RNA by using an Iscript cDNA synthesis kit (Bio-Rad). Quantitative real-time PCR to quantify NDRG4 mRNA levels was performed by using SYBR Green PCR master mix (Applied Biosystems, Nieuwekerk a/d IJssel, the Netherlands) as described previously.[22] Cyclophilin A was used as a reference gene for normalization. Primers used are listed in Supplementary Table S4.3.

Immunohistochemistry

Immunohistochemistry was performed on formalin-fixed, paraffin-embedded tissue sections (5 μm thick). Sections were deparaffinized in xylene, rehydrated, and incubated with 0.3% hydrogen peroxide in methanol for 30 minutes. The sections were incubated with Tris-buffered saline (TBS), 20% fetal calf serum (FCS), and 0.1% Tween to block nonspecific antibody binding, followed by incubation with the anti-NDRG4 monoclonal antibody (Abnova Corporation) diluted 1:6000 in TBS with 0.1% Tween and

0.5% bovine serum albumin (BSA). Sections were incubated with a horseradish peroxidase–conjugated secondary antibody against mouse, rabbit, and rat IgGs (Poly-HRP GAM/R/R IgG) (Immunovision Technologies, Burlingame, CA), and bound antibody was visualized by using 3,3-diaminobenzidine (DAB) substrate as a chromogen (Dako, Glostrup, Denmark) followed by hematoxylin counterstaining.

Colony Formation Assay

Colorectal cancer RKO and HCT116 cells were transfected in six-well plates (1×10^6 cells per well) with pCMV6 or pCMV6-NDRG4 as described above. The next day, the cells were diluted 1:20 and G418 (at 1 mg/ml for RKO 400 μg/ml for HCT116) was added to the medium to select for cells in which the plasmids had stably integrated into genomic DNA. After 14 days of selection, colonies were stained by using Giemsa's azur eosin methylene blue solution (Merck) and counted. Colony formation was assessed in four different experiments (two replicate wells per experiment).

In vitro cell proliferation, migration, and invasion assays

HCT116 cells were seeded onto 96-well plates (5000 cells per well) and cell numbers were counted 24, 48, 72, and 96 hours later (three wells per time point). In addition, after 96 hours of incubation, the cultures were pulse labeled for 6 hours with [methyl-^{3}H] thymidine (0.3 μCi per well; Amersham Life Science, Roosendaal, the Netherlands). Cells were harvested by using a cell harvester and [^{3}H]-thymidine activity was measured by using a liquid scintillation counter. Three independent experiments were performed (three replicate wells per experiment).

Cell migration and invasion assays were performed using Matrigel-coated (invasion assay) or uncoated (migration assay) 24-well transwell plates (8-μm pore size) (BD Biosciences, Franklin Lakes, NJ). Briefly, $2\text{x}10^5$ HCT116 cells in DMEM containing 1% FCS were seeded into the upper chamber of each well, and DMEM containing 20% FCS was placed in the lower chamber. After 48 hours of incubation, the transwells were disassembled and the membranes that separated the upper and lower chamber of each transwell were fixed with methanol and stained with 1% Toluine Blue in 1% borax and the cells on the lower surface of the membrane were counted with the use of a light microscope. Transwell experiments were assessed in three different experiments (two replicate wells per experiment).

Statistical Analysis

For comparison between *NDRG4* methylation frequencies in normal, adenoma and carcinoma tissues from colorectal cancer patients and normal and adenoma tissues from control subjects without colorectal cancer, we used logistic regression (Table 4.1).

Because we observed statistically significant differences in age between the cancer patients and control subjects (analyzed using one-way ANOVA analysis of variance) and in tumor location among the cancer patients (analyzed using Pearson chi-square test) (Supplementary Table S4.1), logistic regression analyses were adjusted for age and location. To compare the prevalence of *NDRG4* promoter methylation in colorectal cancer tissue in relation to clinicopathological features, the Pearson chi-square test (TNM stage, tumor location and sex) or Fisher exact test (age at diagnoses) was used (Table 4.2). In the hospital-based series, non-cancerous control tissues, adenoma tissues, and carcinoma tissues were obtained from the same patients. These paired samples were analyzed by using the McNemar test to compare *NDRG4* methylation frequencies in carcinoma tissue, adenoma and normal tissue from colorectal cancer patients (Table 4.3). Where appropriate, the Bonferroni method was used to correct for multiple comparisons.

For quantitative methylation-specific PCR analysis, we used receiver operating characteristic (ROC) curve analysis and the area under the curve (AUC) to determine the best cutoff value for highest sensitivity and specificity. *NDRG4* promoter methylation was considered positive if the methylation value was greater than the cutoff. Because age differences were expected between cancer patients and control subjects, a ROC curve and generalized linear (ROC-GLM) regression model was used to assess the influence of the age difference on the accuracy of *NDRG4* methylation as a biomarker for the detection of colorectal cancer.

Analysis of cell growth curves was performed by means of two-way ANOVA. The Student *t* test was used for analyses of ^{3}H-thymidine incorporation and anchorage-independent cell growth. The Mann-Whitney rank sum test was used to analyze data obtained in the colony formation, quantitative real-time RT-PCR, migration, and invasion assays.

All *p* values are two-sided, and *P* values less than or equal to .05 were considered statistically significant. Data analysis was done by using SPSS software (version 12.0.1; SPSS Inc., Chicago, IL).

Results

NDRG4 promoter methylation and mRNA expression in colorectal cancer cell lines

The promoter region of the *NDRG4* gene (National Center for Biotechnology Information [NCBI] accession number NM_020465) contains a dense CpG island located from nucleotides -556 to +869 relative to the transcription start site (Figure 4.1A). To assay this region for potential methylation, we examined eight human colorectal cancer cell lines by methylation-specific PCR using primers located from -250 to +10 relative to the transcription start site (primers are listed in Supplementary Table S4.3). The *NDRG4* promoter was methylated in all of the cell lines except SW480 (Figure 4.1B). To

investigate the pattern of CpG island methylation in the *NDRG4* promoter, we sequenced sodium bisulfite-modified genomic DNA isolated from HCT116 (*NDRG4* promoter methylation positive) and SW480 (*NDRG4* promoter methylation negative) cells. The promoter region spanning 39 CpG sites (-251 to +10) was PCR-amplified using sodium bisulfite-modified genomic DNA as template. Bisulfite sequencing confirmed the methylation-specific PCR data, in that HCT116 cells showed almost complete methylation, whereas SW480 cells showed almost no methylated CpGs (Figure 4.1C). To investigate whether promoter methylation was associated with inhibition of gene expression, we measured NDRG4 mRNA levels in HCT116 and RKO cells incubated with and without the DNA methylation inhibitor DAC. In both cell lines, endogenous NDRG4 mRNA levels were statistically significantly higher in DAC-treated cells than in untreated cells (DAC treated vs untreated, RKO cells: 4.4-fold increase, 95% confidence interval [CI] = 3.17- to 5.63-fold increase, p=.014; HCT116 cells: 1.7-fold increase, 95% CI = 0.93- to 2.47-fold increase, p=.037) (Figure 4.1D).

Prevalence of NDRG4 promoter methylation in primary colorectal adenomas and carcinomas

Sequence analysis of sodium bisulfite–modified genomic DNA isolated from three pairs of primary colorectal cancer tissues and matched noncancerous normal colon mucosa showed dense methylation at the *NDRG4* promoter (region –251 to +10 relative to the transcription start site) in the colorectal cancers but almost no *NDRG4* promoter methylation in matched normal colon mucosa (Figure 4.2A). It is interesting that the density of methylation was higher in the upstream region (i.e., the region more 5' relative to the transcription start site) of the *NDRG4* CpG island than in the more downstream region (Figure 4.2A).

To examine the methylation status of the *NDRG4* promoter in a large series of noncancerous colon mucosa, adenoma, and colorectal cancer tissue, we performed methylation-specific PCR with primer pair 1 (which is situated more to the 5'-end of the *NDRG4* promoter region, where the *NDRG4* methylation density was higher compared with the upstream region, Figure 4.1A). The frequency of *NDRG4* promoter methylation was lower in the normal mucosa from the control subjects than in the colorectal cancer tissue from the colorectal cancer patients (2/48 [4%] vs. 71/83 [86%]; p<.001) (Table 4.1). The frequency of *NDRG4* promoter methylation in adjacent normal mucosa of colorectal cancer patients did not differ statistically significantly from that in the normal mucosa of control subjects (9/78 [12%] vs. 2/48 [4%]; p>.99) (Table 4.1). Little or no *NDRG4* promoter methylation was found in skin, renal cell, ovarian, prostate, breast, or esophageal squamous cell carcinomas (data not shown).

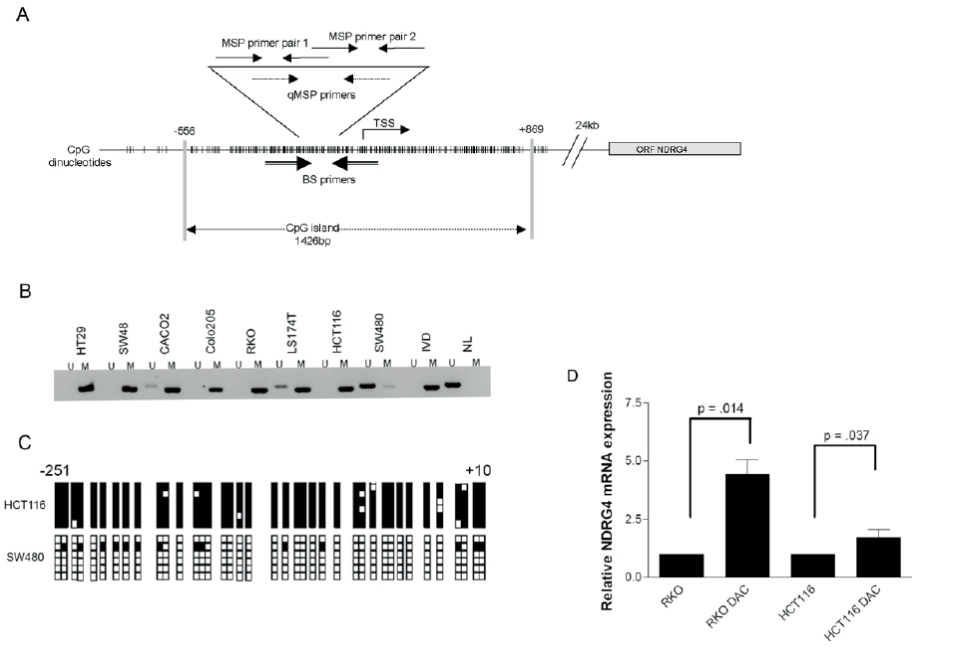

Figure 4.1 NDRG4 promoter structure, promoter methylation, and mRNA expression in colorectal cancer cell lines.

A) Schematic representation of the promoter region of NDRG4 (NM_020465). A dense CpG island is located between nucleotides −556 and +869 relative to the transcription start site (TSS). *Vertical lines* represent the locations of CpG dinucleotides, the *gray rectangle* indicates the open reading frame (ORF) of NDRG4, and *paired arrows* indicate the locations of the amplicons identified by methylation-specific polymerase chain reaction (MSP), quantitative MSP (qMSP), and bisulfite sequencing (BS) primers. **B**) Electrophoretic analysis of MSP amplification products in eight colorectal cancer cell lines. *U* = unmethylated, *M* = methylated, *IVD* = in vitro methylated DNA, *NL* = normal lymphocytes. **C**) Bisulfite sequencing of colorectal cancer HCT116 and SW480s. Six different bacterial clones were sequenced. Each *row* represents an individual cloned allele that was sequenced following sodium bisulfite DNA modification. Each box indicate a CpG dinucleotide (*black box* = methylated CpG site; *white box* = unmethylated CpG site). **D**) NDRG4 mRNA expression in colorectal cancer RKO and HCT116 cells with and without treatment with the DNA methylation inhibitor 5-aza-2'-doxycytidine (DAC). Quantification is presented as mean values (*error bars* correspond to 95% confidence intervals) relative to untreated cells from three independent experiments (normalization was against cyclophillin A). Statistical analysis was done using the Mann–Whitney rank sum test (two-sided).

By contrast, the *NDRG4* promoter was frequently methylated in adenocarcinomas of the esophagus (13/16 [81%]) and in diffuse-type (8/11 [73%]) and intestinal-type (9/11 [82%]) adenocarcinomas of the stomach (data not shown). We also compared the frequency of *NDRG4* promoter methylation in premalignant lesions from patients with and without colorectal cancer. We observed no statistically significant difference in the frequency of *NDRG4* promoter methylation between adenomas obtained from colorectal cancer patients that developed synchronously or metachronously to the

tumor and adenomas obtained from patients that did not develop colorectal cancer after 10 years of follow-up (41/62 [66%] vs 12/22 [55%]; p>.99) (Table 4.1).

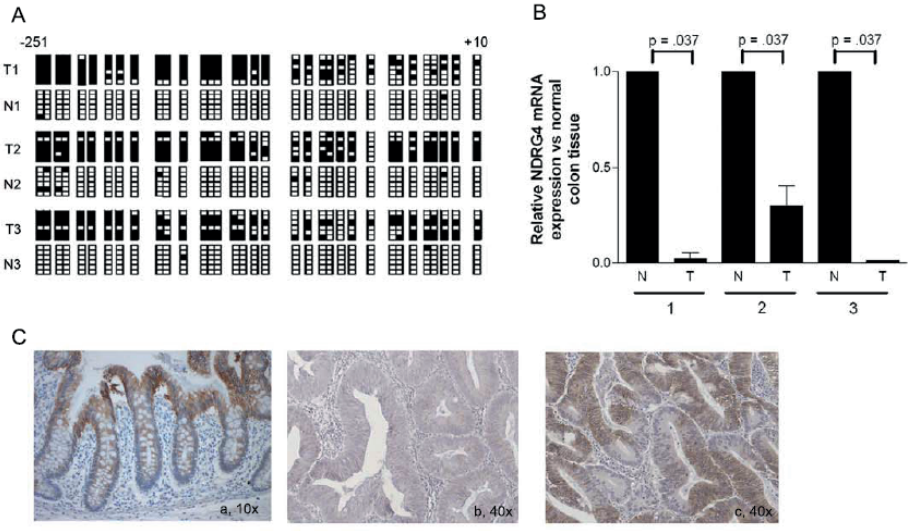

Figure 4.2 ***NDRG4* promoter methylation and expression in primary colorectal cancer.**
A) Bisulfite sequencing of colorectal cancer (T1–3) and the matched normal colon mucosa (N1-3) from three colorectal cancer patients. Six different bacterial clones per tissue sample were sequenced. Each *row* represents an individual cloned allele that was sequenced following sodium bisulfite DNA modification. Each box indicates a CpG dinucleotide (*black box* = methylated CpG site; *white box* = unmethylated CpG site). **B**) NDRG4 mRNA levels measured by real-time polymerase chain reaction in colon cancer tissues (T) and matched normal colon tissue samples (N) from the three colorectal cancer patients in (A). For each patient, the level of NDRG4 mRNA expression in normal mucosa tissue was set to equal 1. Data are presented as mean values relative to normal mucosa tissue of three independent experiments (*error bars* correspond to 95% confidence intervals). Statistical analysis was performed using the Mann–Whitney rank sum test (two-sided). **C**) Immunohistochemical localization of NDRG4 protein expression in normal colon mucosa (a, 10x magnification) and colorectal cancer (b,c, 40x magnification). *Blue* (hematoxylin) staining represents the nuclear staining and *brown* (DAB) staining represents the antibody (NDRG4). Each image is from a different patient.

To confirm the high prevalence of *NDRG4* promoter methylation in colorectal cancer, we analyzed a second independent, population-based series of colorectal cancers and observed that *NDRG4* promoter methylation was present in 70% (128/184) of colorectal cancer patients (data not shown). In each of the two independent series of colorectal cancer patients, *NDRG4* promoter methylation was not associated with age at diagnosis (mean age = 70 years), sex, proximal versus distal tumor location, or TNM stage (Table 4.2).

Table 4.1 ***NDRG4*** **promoter methylation frequencies in normal, adenoma, and carcinoma tissue from colorectal cancer (CRC) patients and normal and adenoma tissue from control subjects without CRC.***

Methylation-specific PCR primer	Carcinoma from CRC patients	Normal tissue from control subjects	p	Normal tissue		p	Adenoma tissue		p
				CRC patients	Control subjects		CRC patients	Control subjects	
NDRG4 p1	71/83 (86)	2/48 (4)	<.001	9/78 (12)	2/48 (4)	>.99	41/62(66%)	12/22 (55)	>.99
NDRG4 p2	55/77 (71)	0/28 (0)	<.001	2/80 (3)	0/28 (0)	>.99	24/58 (41%)	4/31 (13)	>.99

*Methylation frequency is presented as the number of methylated samples divided by the total number of samples analyzed (%). P values (two-sided) were generated by using logistic regression with adjustment for age (continuous data) and tumor location (categories: proximal and distal location). P values are Bonferroni adjusted, and the cutoff for statistical significance is p=.017. PCR = polymerase chain reaction; p1 = methylation-specific primer pair 1; p2 = methylation-specific primer pair 2.

Table 4.2 **Prevalence of** ***NDRG4*** **promoter methylation in colorectal cancer tissue in relation to clinicopathological features for two independent series.***

Characteristic	Hospital-based series†	Population-based series‡
TNM stage§		
I	11/12 (92)	30/42 (71)
II	23/28 (82	42/57 (74)
III	29/32 (91)	39/56 (70)
IV	8/11 (72)	17/21 (81)
P	.431	.790
Tumor location		
Proximal	34/39 (87)	47/58 (81)
Distal	37/42 (89)	81/118 (69)
P	1.00	.141
Sex		
Male	34/41 (83)	71/95 (75)
Female	37/42 (88)	57/81 (70)
P	.548	.611
Age at diagnosis		
≤70 y	30/32 (94)	83/117 (71)
>70 y	41/51 (80)	45/59 (76)
P	.117	.453

* Methylation frequency is presented as the number of methylated samples divided by the total number of samples analyzed (%). *P* values are from two-sided chi-square tests. TNM = tumor–node–metastasis. † Collected from the tissue archive of the department of Pathology of the University Hospital Maastricht. ‡ Prospective Netherlands Cohort Study on Diet and Cancer. § Reference 23.

We next investigated whether *NDRG4* promoter methylation changes during colorectal cancer progression by comparing the frequency of *NDRG4* promoter methylation in samples of normal mucosa, adenoma, and colorectal cancer tissue (Table 4.3). The *NDRG4* promoter was more frequently methylated in colorectal carcinomas than in matched normal mucosa adjacent to the tumor (26/31 [84%] versus 5/31 [16%]; p<.001) (Table 4.3). Adenoma samples from colorectal cancer patients also had a statistically significantly higher *NDRG4* promoter methylation frequency than

normal colon samples (22/36 [61%] vs 5/36 [14%], p<.001) (Table 4.3). Finally, the frequency of *NDRG4* promoter methylation was higher in colorectal carcinomas than in matched adenoma samples, but the difference was not statistically significant (26/32 [81%] vs. 20/32 [63%]; p=.54) (Table 4.3).

Heterogeneity of NDRG4 promoter methylation

As described above, we observed that the density of methylation was higher in the upstream region of the *NDRG4* promoter CpG island than in the more downstream region. We therefore used methylation-specific PCR primer pair 2, which amplifies a region downstream of primer pair 1 (Figure 4.1A), to investigate this region for *NDRG4* promoter methylation.

Table 4.3 ***NDRG4*** **promoter methylation frequencies in carcinoma, adenoma and normal tissue from colorectal cancer patients.***

Methylation-specific PCR primer	Normal tissue	Adenoma tissue	p	Normal tissue	Carcinoma tissue	p	Adenoma tissue	Carcinoma tissue	p
NDRG4 p1	5/36 (14)	22/36 (61)	<.001	5/31 (16)	26/31 (84)	<.001	20/32 (63)	26/32 (81)	.540
NDRG4 p2	0/32 (0)	11/32 (34)	.003	0/30 (0)	22/30 (73)	<.001	13/33 (39)	25/33 (76)	.012

*Methylation frequency is presented as the number of methylated samples divided by the total number of samples analyzed (%). Frequencies for some types of tissue vary because for some patients we did not have simultaneous information on both types of tissue in the analysis (e.g., not all patients with adenomas also had both normal and carcinoma tissue available). *P* values (Bonferroni adjusted) are from two-sided the McNemar test; the cutoff for statistical significance is *P* = .017.

The sensitivity for colorectal cancers (the proportion of people with disease who have a positive test result) decreased from 86% with primer pair 1 to 71% with primer pair 2 while the specificity (the proportion of people without disease with a negative test result) increased from 96% to 100%. Intriguingly, using primer pair 2, we found a statistically significant difference in the frequency of *NDRG4* promoter methylation between matched adenomas and carcinomas from the colorectal cancer patients (13/33 [39%] versus 25/33 [76%], p=.012), which was not observed using primer pair 1 (Table 4.3).

NDRG4 mRNA and protein expression in colorectal cancer

We next examined whether methylation of the CpG island in the *NDRG4* promoter is associated with gene silencing by investigating NDRG4 mRNA expression in colorectal cancer tissue and matched normal colon mucosa from three colorectal cancer patients. In each of the three pairs of tissue, the NDRG4 mRNA level in the colorectal cancer was statistically significantly lower than that in the matched normal colon mucosa (p=.037 for all three matched pairs). Compared with the NDRG4 mRNA level in the matched normal tissue (set at 100%), the expression of *NDRG4* was in tumor 1, 2, and 3 was 3% (95% CI = −3% to 8%), 31% (95% CI = 9% to 58%), and 1.5% (95% CI = 1.3% to 1.7%), respectively (Figure 4.2B).

Next, we performed immunohistochemistry to investigate NDRG4 protein expression in a matched pair of normal colonic mucosa and colorectal cancer from a different colorectal cancer patient. In normal colon mucosa, NDRG4 expression was predominantly in basolateral membranes within colonocytes and increased in intensity near the mucosal surface (Figure 4.2C, panel a). The matched colorectal cancer showed heterogeneous cytoplasmatic staining with the anti-NDRG4 antibody. Similar results were observed in matched pairs of tissues from other colorectal cancer patients (data not shown). In most tumors, less than 50% of the tumor area showed weak to focally strong *NDRG4* expression (Figure 4.2C, panels b and c).

To investigate the association between *NDRG4* promoter methylation and NDRG4 expression, we performed immunohistochemical analysis of NDRG4 protein expression on tissues from 80 colorectal cancer patients of the population-based series. We observed no association between *NDRG4* promoter methylation and NDRG4 expression, suggesting that an alternative mechanism(s) might account for *NDRG4* inactivation in colorectal cancer. Therefore, we analyzed macrodissected colorectal cancer tissue and matched normal tissues from 86 colorectal cancer case patients of the population-based series for loss of heterozygosity. In addition, 12 primary colorectal cancers and the colorectal cancer cell lines HCT116 and SW480 were analyzed for *NDRG4* mutations. We observed loss of heterozygosity in 27 (31%) of 86 colorectal cancers. No inactivating mutations within the coding region of the *NDRG4* gene were detected in the 12 colorectal carcinomas. However, we found one novel non-synonymous mutation in the SW480 cell line (40662A→AG Ile65Val [an A-to-G substitution at nucleotide 4066, resulting in an isoleucine-to-valine substitution at amino acid 65]). In addition, two previously reported single-nucleotide polymorphisms (SNPs) were detected among the 12 colorectal cancers. One SNP was observed in one of the 12 colorectal cancers (43760G→GG Val224Val [a G-to-GG substitution at nucleotide 43760, no difference in amino acid]); National Center for Biotechnology Information [NCBI] SNP database [dbSNP] accession number rs17821543). The second SNP was observed in nine of the 12 colorectal cancers (48311A→AG Ser354Ser [an A-to-AG substitution at nucleotide 43760, no difference in amino acid]; NCBI dbSNP accession number rs42945).

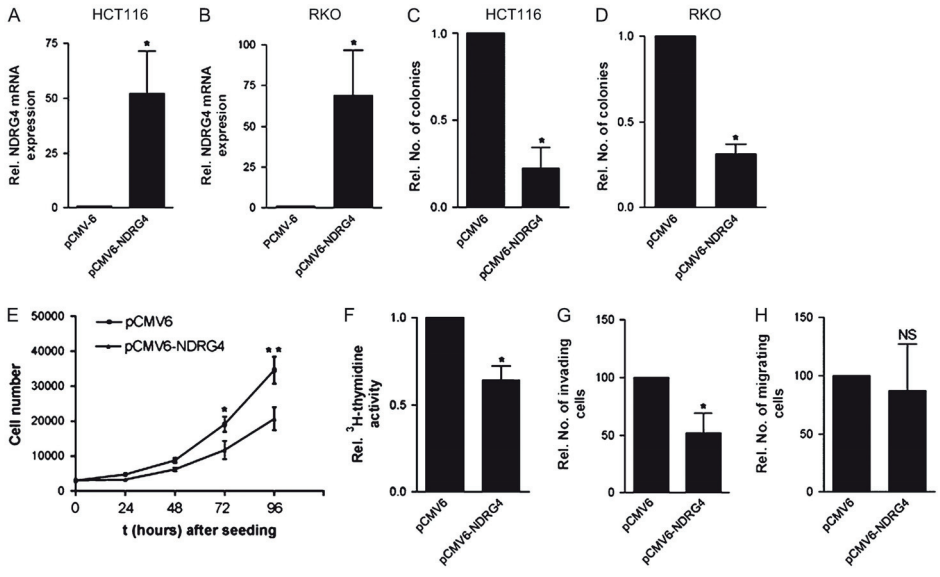

Figure 4.3 **Functional assays of NDRG4 in colorectal cancer.**
(**A**, **B**) NDRG4 mRNA expression measured by real-time reverse transcription–polymerase chain reaction in HCT116 (A) and RKO (B) cells stably transfected with NDRG4 expression vector (pCMV6-NDRG4) or empty vector (pCMV-6). Results are plotted as mean values of mRNA expression relative to control vector in five (HCT116) and three (RKO) independent experiments (HCT116; *P = .005, RKO; *P = .037). (**C**, **D**) Colony formation by HCT116 (C) and RKO (D) cells transfected with pCMV6 or pCMV6-NDRG4 and grown for 2 weeks in medium containing G418. Results are plotted as the mean colony numbers relative to pCMV6 transfectants in four independent experiments (C and D: *P = .014). Statistical analysis in panels A–D was performed using the Mann–Whitney rank sum test (two-sided). **E**) Cell proliferation assay. *NDRG4*-transfected HCT116 cells (pCMV6-NDRG4) were compared with control cells transfected with empty vector (pCMV6). Results are plotted as the mean cell number in three independent experiments (72 hours: *P=.05; 96 hours: **P<.001; two-way analysis of variance). **F**) Cell proliferation measured by ^{3}H-thymidine incorporation. Data are expressed as mean number of proliferating *NDRG4*-transfected cells relative to control cells transfected with empty vector (pCMV6). Three independent experiments were performed (three replicate wells per experiment) (*P<.001; two-sided Student *t* test). **G**) Invasion of HCT116 cells through matrigel-coated transwells. Results represent mean number of *NDRG4*-transfected cells that passed through the matrigel-coated membranes of the transwell relative to control cells transfected with empty vector in three independent experiments (*P<.001, two-sided Mann–Whitney rank sum test). **H**) Migration assay. Plotted are the mean number of *NDRG4*-transfected HCT116 cells that migrated through transwell membranes not coated with matrigel relative to control cells transfected with empty vector in three independent experiments (NS=not statistically significantly different from control, Mann–Whitney rank sum test). Error bars correspond to 95% confidence intervals.

Effect of NDRG4 overexpression in colorectal cancer cell lines

To examine whether NDRG4 acts as a tumor suppressor in colorectal cancer cells, we characterized HCT116 and RKO cells that were transfected with an expression vector harboring the full-length *NDRG4* cDNA or empty vector (control). *NDRG4* transfectants (pools of stably transfected cells) showed increased expression of NDRG4 mRNA compared with control cells transfected with empty vector (HCT116 52%, p=.005; RKO 69%, p=.037). In addition, NDRG4 protein expression was increased in the *NDRG4* transfectants compared with the control cells transfected with empty vector (data not shown). Compared with control transfectants, transfection with *NDRG4* statistically significantly reduced the number of G418-resistant colonies (HCT116: 77% reduction [95% CI = 66% to 90% reduction], p=.014; RKO: 69% reduction [95% CI = 61% to 76% reduction], p=.014) (Figure 4.3C and 4.3D) and statistically significantly decreased cell proliferation as measured by ^{3}H-thymidine activity (HCT116: 40% reduction [95% CI = 27% to 44% reduction], p<.001) (Figure 4.3E and 4.3F). *NDRG4* transfectants displayed reduced invasion through matrigel-coated transwell membranes compared with control transfectants (HCT116: 48% reduction [95% CI = 31% to 64% reduction], p<.001) (Figure 4.3G). However, transfection of *NDGR4* had no effect on HCT116 cell migration compared with control-transfected cells (Figure 4.3H). Taken together, these data suggest that NDRG4 exhibits tumor suppressive effects in human colorectal cancer cells.

Sensitivity and specificity of NDRG4 promoter Methylation in Fecal DNA for the Detection of Colorectal Cancer

The high prevalence of *NDRG4* promoter methylation in colorectal cancer and the absence of methylation in normal colon mucosa suggested that *NDRG4* promoter methylation could be a sensitive and specific biomarker for the noninvasive detection of colorectal cancer in human stool. Therefore, we developed a quantitative molecular beacon–based methylation-specific PCR assay that used a primer pair situated between MSP primer pair 1 and MSP primer pair 2 and fecal DNA isolated from stool (Figure 4.1A). We first examined *NDRG4* promoter methylation as a biomarker for colorectal cancer in a training set comprising 28 colorectal cancer patients and 45 healthy control subjects. We used the data from the training set to construct a receiver operating characteristic (ROC) curve with an AUC of 0.77 (95% CI = 0.66 to 0.86) (Figure 4.4). Using a cutoff level of 1.22 copies, which gave the highest sensitivity and specificity for the detection of colorectal cancer, we detected *NDGR4* promoter methylation in 17 of the 28 colorectal cancer patients, yielding a sensitivity of 61% (95% CI = 43% to 79%) for the detection of colorectal cancer. Three (7%) of the 45 healthy control subjects tested positive for *NDRG4* methylation, which resulted in a specificity of the assay of 93% (95% CI = 90% to 97%). To test the accuracy of the *NDRG4* promoter methylation cutoff generated from the training set, we assayed *NDRG4* promoter methylation by using

quantitative methylation-specific PCR in an independent series of 47 colorectal cancer patients and 30 healthy control subjects, which were handled in a blinded fashion. Using the previously determined cutoff level of 1.22 methylated copies, we found that 25 of the 47 colorectal cancer patients tested positive for *NDRG4* methylation, resulting in a sensitivity of 53% (95% CI = 39% to 67%), and that none of the 30 control subjects tested positive, yielding a specificity of 100% (95% CI = 86% to 100%).

Because the mean age of the colorectal cancer patients differed statistically significantly from that of the control subjects in the training and test set (Supplementary Table S4.4), we used a ROC-GLM regression model in the training set to assess the accuracy of *NDRG4* promoter methylation for the detection of colorectal cancer after adjustment for age. This analysis indicated that age did not statistically significantly influence the accuracy (p=.89, ROC-GLM regression model)[23]. These data indicate that detection of *NDRG4* promoter methylation in fecal DNA can be used as a novel biomarker for detection of colorectal cancer.

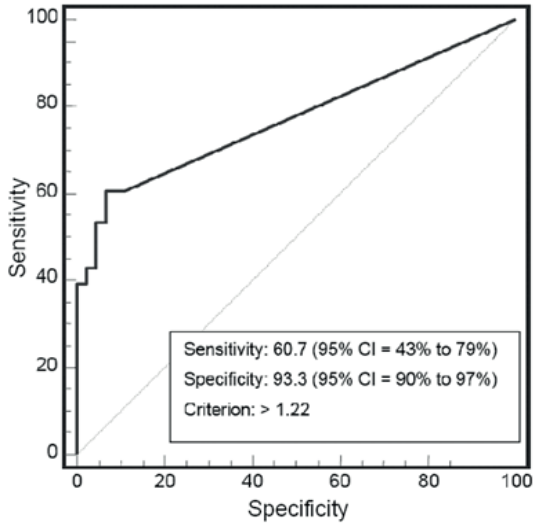

Figure 4.4 ***NDRG4* promoter methylation in fecal DNA as biomarker for colorectal cancer detection.** Sensitivity and specificity at various cutoff values for the training set (which consisted of 28 colorectal cancer patients and 45 healthy control subjects) to obtain a positive test for *NDRG4* Q-MSP are shown in the receiver operating characteristic (ROC) curve. Sensitivity and specificity at various cutoff values for the training set to obtain a positive test for *NDRG4* quanitative methylation-specific polymerase chain reaction are shown. The *jagged line* represents the ROC curve. The *dashed line* represents the line of no discrimination between good and bad classification. The determined optimal cutoff value for *NDRG4* promoter methylation is 1.22 copies.

Discussion

Here we describe the identification and validation of *NDRG4* promoter methylation, a novel, sensitive, and specific marker for the detection of colorectal cancer. We found a statistically significant difference in the frequency of *NDRG4* promoter methylation between colorectal cancers and normal colon mucosa. Adding inflamed mucosa, which often shows promoter CpG island methylation[24], to the normal colon mucosa control group only slightly reduced the specificity of *NDRG4* promoter methylation from 96% to 94%, indicating that *NDRG4* promoter methylation is not associated with inflammation. Although most CpG island promoter methylated genes are frequently associated with a proximal tumor location[25], *NDRG4* promoter methylation is present in both distal colorectal cancers (often associated with chromosomal instability)[26] as well as proximal colorectal cancers (often associated with microsatellite instability). This finding makes *NDRG4* promoter methylation a promising marker to detect chromosomal-instable as well as microsatellite-instable colorectal cancers. We also found that DNA methylation density was higher in the upstream region of the *NDRG4* promoter than in the downstream region. Therefore, we used two different pairs of methylation-specific PCR primers that amplify overlapping fragments in the CpG island to detect *NDRG4* promoter methylation. Using primer pair 1 (the one most 5' to the transcription start site) to detect *NDRG4* promoter methylation resulted in a sensitivity for colorectal cancer of 86% and for adenoma of 66% and a specificity of 96%. Using primer pair 2, the sensitivity for colorectal cancers and adenomas decreased to 71% and 41%, respectively, while the specificity increased to 100%. This finding suggests that *NDRG4* promoter hypermethylation initially occurs at the 5' end of the *NDRG4* CpG island and spreads toward the transcription start site before ultimately shutting down *NDRG4* mRNA expression, as has also been observed in the promoter of for *RUNX3.*[27 28]

In addition, we found a statistically significant lower methylation frequency in adenoma tissue compared with carcinoma tissue when we used primer pair 2 to detect *NDRG4* promoter methylation. This finding suggests that DNA methylation in the promoter of *NDRG4* may spread toward the transcription start site during cancer progression.

To our knowledge, expression of *NDRG4* has been documented only in brain and heart tissue by northern blotting. Here we show that NDRG4 mRNA and protein is expressed in normal colon tissue and that expression of both is decreased in colorectal cancer. However, we found no statistically significant association between *NDRG4* promoter methylation and NDRG4 expression. We and others[29] also found no evidence for mutational inactivation of *NDRG4.* However, we observed loss of heterozygosity of the *NDRG4* locus at chromosome 16q in 31% of the colorectal cancers analyzed. Frequent loss of heterozygosity at 16q was previously observed in a wide variety of solid tumor types, including breast[30], liver[31,32], prostate[33], ovarian[34], and Wilms tumors[35], but, to our knowledge, has not been described in colorectal cancer. These findings

suggest that promoter methylation and loss of heterozygosity contribute to the altered expression of *NDRG4* in colorectal cancer.

We also provide evidence that NDRG4 has tumor suppressor activities in colorectal cancer. Overexpression of *NDRG4* in the colorectal cancer HCT116 cells inhibited colony formation and cell proliferation and invasion in vitro, suggesting that *NDRG4* is a tumor suppressor gene in colorectal cancer. The ability of the *NDRG4*-transfected cells to migrate in vitro was identical to that of control cells, which was also observed for *NDRG1*[4].

To investigate the potential utility of *NDRG4* promoter methylation as a noninvasive biomarker test to identify individuals who should under go colonoscopy because they are at increased risk for colorectal cancer, we analyzed fecal DNA of colorectal cancer patients and control subjects. We showed that detection of *NDRG4* promoter methylation in fecal DNA was both sensitive and specific at identifying colorectal cancer patients in two independent series of case patients and control subjects. The prevalence of *NDRG4* promoter methylation in colorectal cancer tissue (86% and 70% in two independent series) suggests that the use of optimal protocols for the isolation of fecal DNA could, in theory, yield a stool *NDRG4* quantitative methylation-specific PCR test with an even higher sensitivity. Methyl-binding domain protein columns to capture methylated DNA, which have been shown to markedly increase sensitivity without decreasing specificity[36], may be useful in this respect. Several studies have provided proof of principle for the detection of promoter CpG island hypermethylation of colorectal adenoma or colorectal cancer–derived DNA in stool[37-43] and blood[44-49]. Compared with other gene promoter methylation markers described thus far[37,39,44-48], *NDRG4* promoter methylation performs well as a novel single marker. Combining *NDRG4* promoter methylation with other DNA markers could improve its sensitivity and specificity for the detection of colorectal cancer, as has been observed for other methylation markers.[50,51] In addition, the specificity of *NDRG4* promoter methylation for gastrointestinal adenocarcinomas makes it a specific biomarker for detecting gastrointestinal cancers in stool and blood.

The stool study presented here should be considered a pilot study that has specific limitations. Not all stool samples from colorectal cancer patients were collected before colonoscopy as was done for the control subjects. It is therefore possible, albeit very unlikely, that colonoscopy could have introduced artifacts in the observed methylation prevalence. In addition, the age of the colorectal cancer patients and control subjects differed statistically significantly, which raises the question of whether the detected *NDRG4* promoter methylation is associated with age. However, this possibility is highly unlikely because neither data from the two independent colorectal cancer series nor the ROC-GLM regression analysis of the stool samples showed that *NDRG4* promoter methylation was associated with age. Nevertheless, the stool data should be validated in large prospective screening study for colorectal cancer, as should

NDRG4 promoter methylation be combined with other methylation markers to enhance the sensitivity and/or specificity.

In conclusion, to our knowledge, this is the first study to describe a tumor suppressor role for *NDRG4* in cancer. Our data indicate that *NDRG4* promoter methylation is potentially useful as a sensitive and specific noninvasive preselection modality for identifying individuals at risk for colorectal cancer for whom colonoscopy is recommended.

References

1. Hellebrekers DM, Melotte V, Vire E, Langenkamp E, Molema G, Fuks F, Herman JG, Van Criekinge W, Griffioen AW, van Engeland M. Identification of epigenetically silenced genes in tumor endothelial cells. *Cancer Res* 2007;67(9):4138-4148.
2. Zhou RH, Kokame K, Tsukamoto Y, Yutani C, Kato H, Miyata T. Characterization of the human NDRG gene family: a newly identified member, NDRG4, is specifically expressed in brain and heart. *Genomics* 2001;73(1):86-97.
3. Qu X, Zhai Y, Wei H, Zhang C, Xing G, Yu Y, He F. Characterization and expression of three novel differentiation-related genes belong to the human NDRG gene family. *Mol Cell Biochem* 2002;229 (1-2):35-44.
4. Bandyopadhyay S, Pai SK, Gross SC, Hirota S, Hosobe S, Miura K, Saito K, Commes T, Hayashi S, Watabe M, Watabe K. The Drg-1 gene suppresses tumor metastasis in prostate cancer. *Cancer Res* 2003; 63(8):1731-1736.
5. Bandyopadhyay S, Pai SK, Hirota S, Hosobe S, Takano Y, Saito K, Piquemal D, Commes T, Watabe M, Gross SC, Wang Y, Ran S, Watabe K. Role of the putative tumor metastasis suppressor gene Drg-1 in breast cancer progression. *Oncogene* 2004;23(33):5675-5681.
6. Guan RJ, Ford HL, Fu Y, Li Y, Shaw LM, Pardee AB. Drg-1 as a differentiation-related, putative metastatic suppressor gene in human colon cancer. *Cancer Res* 2000;60(3):749-755.
7. van Belzen N, Dinjens WN, Diesveld MP, Groen NA, van der Made AC, Nozawa Y, Vlietstra R, Trapman J, Bosman FT. A novel gene which is up-regulated during colon epithelial cell differentiation and down-regulated in colorectal neoplasms. *Lab Invest* 1997;77(1):85-92.
8. Kurdistani SK, Arizti P, Reimer CL, Sugrue MM, Aaronson SA, Lee SW. Inhibition of tumor cell growth by RTP/rit42 and its responsiveness to p53 and DNA damage. *Cancer Res* 1998;58(19):4439-4444.
9. Shah MA, Kemeny N, Hummer A, Drobnjak M, Motwani M, Cordon-Cardo C, Gonen M, Schwartz GK. Drg1 expression in 131 colorectal liver metastases: correlation with clinical variables and patient outcomes. *Clin Cancer Res* 2005;11(9):3296-3302.
10. Lusis EA, Watson MA, Chicoine MR, Lyman M, Roerig P, Reifenberger G, Gutmann DH, Perry A. Integrative genomic analysis identifies NDRG2 as a candidate tumor suppressor gene frequently inactivated in clinically aggressive meningioma. *Cancer Res* 2005;65(16):7121-7126.
11. Deng Y, Yao L, Chau L, Ng SS, Peng Y, Liu X, Au WS, Wang J, Li F, Ji S, Han H, Nie X, Li Q, Kung HF, Leung SY, Lin MC. N-Myc downstream-regulated gene 2 (NDRG2) inhibits glioblastoma cell proliferation. *Int J Cancer* 2003;106(3):342-347.
12. Liu N, Wang L, Liu X, Yang Q, Zhang J, Zhang W, Wu Y, Shen L, Zhang Y, Yang A, Han H, Zhang J, Yao L. Promoter methylation, mutation, and genomic deletion are involved in the decreased NDRG2 expression levels in several cancer cell lines. *Biochem Biophys Res Commun* 2007;358(1):164-169.
13. Nakada N, Hongo S, Ohki T, Maeda A, Takeda M. Molecular characterization of NDRG4/Bdm1 protein isoforms that are differentially regulated during rat brain development. *Brain Res Dev Brain Res* 2002;135(1-2):45-53.
14. Ohki T, Hongo S, Nakada N, Maeda A, Takeda M. Inhibition of neurite outgrowth by reduced level of NDRG4 protein in antisense transfected PC12 cells. *Brain Res Dev Brain Res* 2002;135(1-2):55-63.
15. Maeda A, Hongo S, Miyazaki A. Genomic organization, expression, and comparative analysis of noncoding region of the rat Ndrg4 gene. *Gene* 2004;324:149-158.
16. Hongo S, Watanabe T, Takahashi K, Miyazaki A. Ndrg4 enhances NGF-induced ERK activation uncoupled with Elk-1 activation. *J Cell Biochem* 2006;98(1):185-193.
17. van den Brandt PA, Goldbohm RA, van 't Veer P, Volovics A, Hermus RJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in The Netherlands. *J Clin Epidemiol* 1990;43(3):285-295.
18. Brink M, de Goeij AF, Weijenberg MP, Roemen GM, Lentjes MH, Pachen MM, Smits KM, de Bruine AP, Goldbohm RA, van den Brandt PA. K-ras oncogene mutations in sporadic colorectal cancer in The Netherlands Cohort Study. *Carcinogenesis* 2003;24(4):703-710.
19. Hermanek P, Scheibe O, Spiessl B, Wagner G. [TNM classification of malignant tumors: the new 1987 edition]. *Rontgenblatter* 1987;40(6):200.

20. Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A* 1996;93(18):9821-9826.
21. van Engeland M, Weijenberg MP, Roemen GM, Brink M, de Bruine AP, Goldbohm RA, van den Brandt PA, Baylin SB, de Goeij AF, Herman JG. Effects of dietary folate and alcohol intake on promoter methylation in sporadic colorectal cancer: the Netherlands cohort study on diet and cancer. *Cancer Res* 2003;63(12):3133-3137.
22. Thijssen VL, Brandwijk RJ, Dings RP, Griffioen AW. Angiogenesis gene expression profiling in xenograft models to study cellular interactions. *Exp Cell Res* 2004;299(2):286-293.
23. Janes H, Pepe MS. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting. *Am J Epidemiol* 2008;168(1):89-97.
24. Issa JP, Ahuja N, Toyota M, Bronner MP, Brentnall TA. Accelerated age-related CpG island methylation in ulcerative colitis. *Cancer Res* 2001;61(9):3573-3577.
25. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, Kang GH, Widschwendter M, Weener D, Buchanan D, Koh H, Simms L, Barker M, Leggett B, Levine J, Kim M, French AJ, Thibodeau SN, Jass J, Haile R, Laird PW. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* 2006;38(7):787-793.
26. Derks S, Postma C, Carvalho B, van den Bosch SM, Moerkerk PT, Herman JG, Weijenberg MP, de Bruine AP, Meijer GA, van Engeland M. Integrated analysis of chromosomal, microsatellite and epigenetic instability in colorectal cancer identifies specific associations between promoter methylation of pivotal tumour suppressor and DNA repair genes and specific chromosomal alterations. *Carcinogenesis* 2008;29(2):434-439.
27. Turker MS. Gene silencing in mammalian cells and the spread of DNA methylation. *Oncogene* 2002;21(35):5388-5393.
28. Homma N, Tamura G, Honda T, Matsumoto Y, Nishizuka S, Kawata S, Motoyama T. Spreading of methylation within RUNX3 CpG island in gastric cancer. *Cancer Sci* 2006;97(1):51-56.
29. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE. The consensus coding sequences of human breast and colorectal cancers. *Science* 2006;314(5797):268-274.
30. Rakha EA, Green AR, Powe DG, Roylance R, Ellis IO. Chromosome 16 tumor-suppressor genes in breast cancer. *Genes Chromosomes Cancer* 2006;45(6):527-535.
31. Bando K, Nagai H, Matsumoto S, Koyama M, Kawamura N, Tajiri T, Onda M, Emi M. Identification of a 1-Mb common region at 16q24.1-24.2 deleted in hepatocellular carcinoma. *Genes Chromosomes Cancer* 2000;28(1):38-44.
32. Sakai K, Nagahara H, Abe K, Obata H. Loss of heterozygosity on chromosome 16 in hepatocellular carcinoma. *J Gastroenterol Hepatol* 1992;7(3):288-292.
33. Elo JP, Harkonen P, Kyllonen AP, Lukkarinen O, Poutanen M, Vihko R, Vihko P. Loss of heterozygosity at 16q24.1-q24.2 is significantly associated with metastatic and aggressive behavior of prostate cancer. *Cancer Res* 1997;57(16):3356-3359.
34. Kawakami M, Staub J, Cliby W, Hartmann L, Smith DI, Shridhar V. Involvement of H-cadherin (CDH13) on 16q in the region of frequent deletion in ovarian cancer. *Int J Oncol* 1999;15(4):715-720.
35. Mason JE, Goodfellow PJ, Grundy PE, Skinner MA. 16q loss of heterozygosity and microsatellite instability in Wilms' tumor. *J Pediatr Surg* 2000;35(6):891-896; discussion 896-897.
36. Zou H, Harrington J, Rego RL, Ahlquist DA. A novel method to capture methylated human DNA from stool: implications for colorectal cancer screening. *Clin Chem* 2007;53(9):1646-1651.
37. Chen WD, Han ZJ, Skoletsky J, Olson J, Sah J, Myeroff L, Platzer P, Lu S, Dawson D, Willis J, Pretlow TP, Lutterbaugh J, Kasturi L, Willson JK, Rao JS, Shuber A, Markowitz SD. Detection in fecal DNA of colon cancer-specific methylation of the nonexpressed vimentin gene. *J Natl Cancer Inst* 2005;97(15):1124-1132.
38. Muller HM, Oberwalder M, Fiegl H, Morandell M, Goebel G, Zitt M, Muhlthaler M, Ofner D, Margreiter R, Widschwendter M. Methylation changes in faecal DNA: a marker for colorectal cancer screening? *Lancet* 2004;363(9417):1283-1285.

39. Lenhard K, Bommer GT, Asutay S, Schauer R, Brabletz T, Goke B, Lamerz R, Kolligs FT. Analysis of promoter methylation in stool: a novel method for the detection of colorectal cancer. *Clin Gastroenterol Hepatol* 2005;3(2):142-149.
40. Leung WK, To KF, Man EP, Chan MW, Bai AH, Hui AJ, Chan FK, Lee JF, Sung JJ. Detection of epigenetic changes in fecal DNA as a molecular screening test for colorectal cancer: a feasibility study. *Clin Chem* 2004;50(11):2179-2182.
41. Belshaw NJ, Elliott GO, Williams EA, Bradburn DM, Mills SJ, Mathers JC, Johnson IT. Use of DNA from human stools to detect aberrant CpG island methylation of genes implicated in colorectal cancer. *Cancer Epidemiol Biomarkers Prev* 2004;13(9):1495-1501.
42. Petko Z, Ghiassi M, Shuber A, Gorham J, Smalley W, Washington MK, Schultenover S, Gautam S, Markowitz SD, Grady WM. Aberrantly methylated CDKN2A, MGMT, and MLH1 in colon polyps and in fecal DNA from patients with colorectal polyps. *Clin Cancer Res* 2005;11(3):1203-1209.
43. Ahlquist DA, Sargent DJ, Loprinzi CL, Levin TR, Rex DK, Ahnen DJ, Knigge K, Lance MP, Burgart LJ, Hamilton SR, Allison JE, Lawson MJ, Devens ME, Harrington JJ, Hillman SL. Stool DNA and occult blood testing for screen detection of colorectal neoplasia. *Ann Intern Med* 2008;149(7):441-450, W481.
44. Ebert MP, Model F, Mooney S, Hale K, Lograsso J, Tonnes-Priddy L, Hoffmann J, Csepregi A, Rocken C, Molnar B, Schulz HU, Malfertheiner P, Lofton-Day C. Aristaless-like homeobox-4 gene methylation is a potential marker for colorectal adenocarcinomas. *Gastroenterology* 2006;131(5):1418-1430.
45. Grady WM, Rajput A, Lutterbaugh JD, Markowitz SD. Detection of aberrantly methylated hMLH1 promoter DNA in the serum of patients with microsatellite unstable colon cancer. *Cancer Res* 2001;61(3):900-902.
46. Leung WK, To KF, Man EP, Chan MW, Bai AH, Hui AJ, Chan FK, Sung JJ. Quantitative detection of promoter hypermethylation in multiple genes in the serum of patients with colorectal cancer. *Am J Gastroenterol* 2005;100(10):2274-2279.
47. Han M, Liew CT, Zhang HW, Chao S, Zheng R, Yip KT, Song ZY, Li HM, Geng XP, Zhu LX, Lin JJ, Marshall KW, Liew CC. Novel blood-based, five-gene biomarker set for the detection of colorectal cancer. *Clin Cancer Res* 2008;14(2):455-460.
48. Zou HZ, Yu BM, Wang ZW, Sun JY, Cang H, Gao F, Li DH, Zhao R, Feng GG, Yi J. Detection of aberrant p16 methylation in the serum of colorectal cancer patients. *Clin Cancer Res* 2002;8(1):188-191.
49. Frattini M, Gallino G, Signoroni S, Balestra D, Battaglia L, Sozzi G, Leo E, Pilotti S, Pierotti MA. Quantitative analysis of plasma DNA in colorectal cancer patients: a novel prognostic tool. *Ann N Y Acad Sci* 2006;1075:185-190.
50. Itzkowitz S, Brand R, Jandorf L, Durkee K, Millholland J, Rabeneck L, Schroy PC, 3rd, Sontag S, Johnson D, Markowitz S, Paszat L, Berger BM. A simplified, noninvasive stool DNA test for colorectal cancer detection. *Am J Gastroenterol* 2008;103(11):2862-2870.
51. Itzkowitz SH, Jandorf L, Brand R, Rabeneck L, Schroy PC, 3rd, Sontag S, Johnson D, Skoletsky J, Durkee K, Markowitz S, Shuber A. Improved fecal DNA test for colorectal cancer screening. *Clin Gastroenterol Hepatol* 2007;5(1):111-117.

Supplementary figure and tables

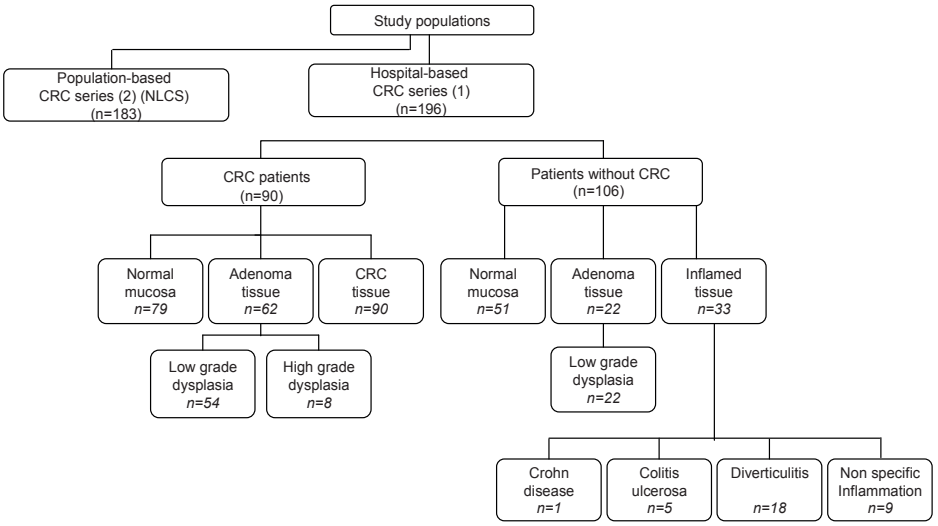

Figure S4.1 Study population.
Of the 90 colorectal cancer patients, carcinoma tissue was collected and, if present, also normal colon mucosa tissue (n=79) and synchronous or metachronous adenoma tissue (n=62). The tissue collected from 106 non-cancerous patients consisted of histologically normal biopsy material (n=51) from patients that underwent endoscopy and did not develop adenomas or colorectal cancer, adenoma biopsy specimens (n=22) from patients who did not develop colorectal cancer within 10 years, and resected colon mucosa from patients with various inflammatory bowel conditions (n=33) who did not develop adenomas or colorectal cancer. CRC = colorectal cancer; NLCS = Netherlands Cohort Study on Diet and Cancer.

Table S4.1 Characteristics of the hospital-based series.*

Subjects	N	Mean age, y (SD)	No. of males/No. of females	Tissue location†	
				No. in proximal colon/Total no. (%)	No. in distal colon/Total no. (%)
Colorectal cancer patients	90				
Normal tissue	79	71.0 (8.6)	41/38	33/75 (44)	42/75 (56)
Adenoma tissue	62	71.7 (7.9)	32/30	20/59 (34)	39/59 (66)
Carcinoma tissue	90	71.5 (8.3)	44/46	41/88 (47)	47/88 (53)
Patients without colorectal cancer	106				
Normal tissue	51	65.2 (9.0)	22/29	9/39 (23)	30/39 (77)
Adenoma tissue	22	63.1 (7.6)	16/6	2/18 (11)	16/18 (89)
Inflamed tissue	33	65.3 (10.1)	14/19	4/26 (15)	22/26 (85)
P		<.001‡	.262§	<.001§	

*Tissue retrospectively collected from the tissue archive of the Department of Pathology of the Maastricht University Medical Center. †Tissue location was not available for all samples. ‡One-way analysis of variance (two-sided). §Pearson chi-square test (two-sided).

Table S4.2 Histological features of colorectal adenoma and carcinoma tissue of the hospital-based series.*

Tissue characteristic	Carcinoma tissue (N=90)	Tissue characteristic	Adenoma tissue	
			Colorectal cancer patients (N=62)	Patients without colorectal cancer (N=22)
Histological type, No. (%)		Histological type, No. (%)		
Adenocarcinoma	72 (80)	Tubular	39 (63)	16 (73)
Mucinous carcinoma	18 (20)	Tubulovillous	22 (36)	6 (27)
		Villous	1 (2)	0 (0)
Differentiation, No. (%)				
Poor	8 (9)	Dysplasia, No. (%)		
Moderate	70 (78)	Low grade	54 (87)	22 (100)
Well	12 (13)	High grade	8 (13)	0 (0)
TNM stage, No. (%)				
I	13 (14)			
II	29 (32)			
III	36 (40)			
IV	12 (13)			

*Tissue retrospectively collected from the tissue archive of the Department of Pathology of the Maastricht University medical center. Percentages in some categories do not total 100% because of rounding. TNM = tumor–node–metastasis.

Table S4.3 Primer sequences and PCR conditions.*

Primers and PCR conditions for nested MSP

NDRG4	Primer	Sense primer (5′→3′)	Antisense primer (5′→3′)	Annealing temp (°C)	No. of PCR cycles
MSP set 1	Flank	ATYGGGGTGTTTTTTAGGTTT	ATACCRAACCTAAAACTAATCCC	56	35
MSP set 1	U	GGGTGTTTTTTAGGTTTTGTGTTGT	CCTAAAACTAATCCCAAACAAACCA	66	30
MSP set 1	M	TTTTTTAGGTTTCGCGTCGC	AAACTAATCCCGAACGAACCG	66	30
MSP set 2	Flank	GGTTYGTTYGGGATTAGTTTTAGG	CRAACAACCAAAAACCCCTC	56	35
MSP set 2	U	GATTAGTTTTAGGTTTGGTATTGTTTTGT	AAAACCAAACTAAAAACAATACACCA	66	25
MSP set 2	M	TTTAGGTTCGGTATCGTTTCGC	CGAACTAAAAACGATACGCCG	66	25

Primers for bisulfite sequencing

Gene	Sense primer (5′→3′)	Antisense primer (5′→3′)
NDRG4	GATYGGGGTGTTTTTTAGGTTT	CRAACAACCAAAAACCCCTC

Primers for quantitative MSP

Gene	Sense primer (5′→3′)	Antisense primer (5′→3′)	Beacon (5′→3′)
Beta-actin	TAGGGAGTATATAGGTTGGGGAAGTT	AACACACAATAACAAACACAAATTCAC	FAM-CGACTGCGTGTGGGGTGGTGATGGAGGAGGTTTAGGCAGTCG-DABCYL
NDRG4	GTATTTTAGTCGCGTAGAAGGC	AATTTAACGAATATAAACGCTCGAC	FAM-CGACATGCCCGAACGAACCGCGATCCCTGCATGTCG-DABCYL

Primers for realtime RT-PCR

Gene	Sense primer (5′→3′)	Antisense primer (5′→3′)
Cyclophilin A	CTCGAATAAGTTTGACTTGTGTTT	CTAGGCATGGGAGGGAACA
NDRG4	GGCCTTCTGCATGTAGTGATCCG	GGTGATCTCCTGCATGTCCTCG

*PCR = polymerase chain reaction; MSP = methylation-specific PCR; U = unmethylated; M = methylated; RT-PCR = reverse transcription–PCR; FAM = 6-carboxyfluorescein.

Table S4.4 Characteristics of subjects who provided stool samples.*

Characteristic	Training set		Test set	
	CRC patients (N=28)	Control patients (N=45)	CRC patients (N=46)	Control patients (N=30)
Mean age, y (SD)	69.29 (9.0)	54.95 (11.5) †	71.13 (9.0)	52.27 (9.8)
Median age, y (IQR)	69.0 (63.3–75.5)	55.0 (52.0–60.0)†	71.5 (64.8–77.3)	53.0 (50.8–57.0)
Male, No. (%)	19 (67.9)‡	19 (42.2)	27 (57.4)§	8 (26.7)
UICC stage, No. (%)				
I	10 (35.7)	–	10 (21.3)	–
II	8 (28.6)	–	19 (40.4)	–
III	8 (28.6)	–	13 (27.7)	–
IV	2 (7.1)	–	4 (8.5)	–
Unknown	0	–	1 (2.1)	–

*CRC = colorectal cancer; IQR = interquartile range; UICC = International Union Against Cancer; – = not applicable. †For one control patient in the training set, age was unknown. ‡For one CRC patient in the training set, sex was unknown. §For one CRC patient in the test set, sex was unknown.

Chapter 5

NDRG4 is predominantly expressed in neurons of the central and peripheral nervous system and restricted to enteric neurons of the mouse and human intestinal tract

Marjolein HFM Lentjes, Nathalie Vaes, Marion J Gijbels, Glenn Rademakers, Kathleen LJ Daenen, Kim AD Wouters, Anneleen Geuzens, Xiang-Hu Qu, Hellen PJ Steinbusch, Bart PF Rutten, Scott H Baldwin, Keith A Sharkey, Robert MW Hofstra, Manon van Engeland, Pieter Vanden Berghe, Veerle Melotte

Submitted

Abstract

Background

N-Myc downstream-regulated gene 4 (*NDRG4)* is a member of the *NDRG* gene family, a group of genes involved in cell proliferation, differentiation, development and stress responses. Promoter methylation of *NDRG4* in fecal DNA has emerged as a promising biomarker for the early detection of colorectal cancer (CRC)**.** Despite its connection to CRC, NDRG4 has been studied predominantly in brain and heart, with little to no knowledge regarding its expression and function in other organs. In this study we investigated the whole-body expression of NDRG4, with a focus on the intestinal tract.

Methods

We determined NDRG4 expression in human and mouse (*NDRG4* wildtype and knockout mice) tissues according to the following specificity criteria: i.e. (1) a correspondent staining pattern using immunohistochemistry and *in situ* mRNA hybridization and (2) absence of immunoreactivity and Western Blot signals in tissues deficient for NDRG4. Additionally, we investigated cell-specific expression of NDRG4 in murine whole-mount preparations in depth using immunofluorescence and confocal microscopy.

Results

We observed that NDRG4 is almost exclusively expressed within nervous system structures throughout the body. In the central nervous system, NDRG4 immunoreactivity is detected in the gray matter of the cerebrum and cerebellum. We observed neuron specific expression within every other organ in the murine body, including the gastro-intestinal tract. In both mouse and human small and large intestines, NDRG4 immunoreactivity is restricted to the enteric nervous system. NDRG4 is expressed in cell bodies of the myenteric and submucosal plexus, nerve fibers connecting the ganglia and nerves innervating the muscularis mucosae and muscularis externa. NDRG4 expression is limited to neurons, as NDRG4 positive cells are always labeled for the pan-neuronal marker HuC/D but never colocalize with the glial cell marker GFAP. Furthermore, NDRG4 expression is observed in a subset of calretinin positive neurons but is virtually absent in the neurons positive for the neuronal nitric oxide synthase.

Conclusion

We conclude that NDRG4 is specifically expressed within the central, peripheral and enteric nervous system. Moreover, we demonstrate that NDRG4 is exclusively expressed by enteric neurons pointing at a neuron-specific role of this protein in the intestinal tract. How the expression pattern of NDRG4 associates with its potential as a biomarker for the early detection of CRC and with its function(s) in the normal and diseased gut, is subject for future investigation.

Introduction

We have identified *NDRG4* promoter CpG island methylation as a promising early detection marker for colorectal cancer (CRC)[1], an observation that has been independently validated by others.[2-5] *NDRG4* has been incorporated into the FDA-approved, multi-target stool DNA test (Cologuard®), which accurately detects CRC and advanced precancerous lesions.[4] Despite its potential as a biomarker, the function of NDRG4 in the gut and its role in CRC carcinogenesis are unknown.

NDRG4, also known as *SMAP-8* and *BDM1*, belongs to the differentiation-related *NDRG* gene family which consists of 4 members, *NDRG1-4*. The NDRG proteins are characterized by an α/β hydrolase fold, share 52-65% sequence homology and are highly conserved during evolution in various species.[6,7] The *NDRG4* gene was originally characterized in humans and encodes three isoforms: NDRG4H (41kDa), NDRG4B (37kDa) and NDRG4B^{var}(39kDa).[6] Even though all three human isoforms have been identified in mouse and rat brain, the latter reveals an additional fourth isoform (45kDa) encoded by the rat ortholog of human *NDRG4* '*SMAP-8/BDM1'*, while the mouse orthologs correspond to NDRG4B and NDRG4B^{var}.[8,9] In contrast to the well-described ubiquitous expression of NDRG1, NDRG2 and NDRG3, abundant expression of NDRG4 is predominantly observed in brain and heart, suggesting an important role of NDRG4 in these organs.[6-11] NDRG4 expression in the central nervous system (CNS) during embryonic zebrafish and mouse development is required for normal morphogenesis of the brain.[8,11] In adult mice, NDRG4 is essential to maintain physiological levels of brain-derived neurotrophic factor (BDNF), which is necessary to retain normal spatial learning/memory and to protect against severe neurological deficits.[8] Furthermore, NDRG4 is required for proper neurite outgrowth, neural survival and function *in vitro* and myelination of axons in zebrafish.[12,13] In line, the Alzheimer diseased brain, which is characterized by neuronal degeneration, shows reduced levels of *NDRG4*.[6] Within the embryonic zebrafish and mouse heart, NDRG4 regulates directional cell movement to ensure proper morphogenesis of the heart and NDRG4 deficiency is associated with weak contractility and reduced heart rate in zebrafish.[11,14] In addition, variations in locus 16q21, which includes the *NDRG4* gene, have been associated with prolongation of the QT interval, a risk factor for sudden cardiac death.[15] The molecular mechanisms by which NDRG4 affects the above described phenomena are still unknown. Furthermore, the expression and functional role(s) of this gene outside the heart and brain are largely unstudied.

In this study we examined the expression of NDRG4 by immunohistochemistry, *in situ mRNA* hybridization and Western blotting using human, *NDRG4* wild-type mice and *NDRG4* knockout mice tissues.[16] We observed NDRG4 expression in numerous neurons of the CNS as previously described. Interestingly, we observed this neuronal-specific NDRG4 immunoreactivity throughout the body, including the heart and the gastro-intestinal tract. This study shows the specific neuronal expression pattern of NDRG4 in

all studied tissues which suggests an important role for NDRG4 in the central, peripheral and enteric nervous system.

Materials and methods

Mice

NDRG4 wild-type (*NDRG4*$^{+/+}$), heterozygous (*NDRG4*$^{+/-}$) and knock-out (*NDRG4*$^{-/-}$) mice, established on a C57BL/6 genetic background as described previously[17] were age- and gender matched and housed in groups of 3 to 5 under standard conditions having free access to food and water. At the age of one year, *NDRG4*$^{+/+}$, *NDRG4*$^{+/-}$ and *NDRG4*$^{-/-}$ mice were sacrificed by CO_2 asphyxiation. Tissues were harvested and either snap-frozen for protein and RNA isolation or fixed in 4% formaldehyde for immunohistochemical analysis. In addition, two month old *NDRG4*$^{+/+}$ and *NDRG4*$^{-/-}$ mice were sacrificed, followed by removal of the intestines for immunofluorescence on whole-mount gut preparations.[18] Animal experiments were approved by the Committee of Animal Welfare of Maastricht University and performed according to Dutch regulations.

Human tissue

Formalin-fixed, paraffin-embedded human normal colon tissues (n=3) were retrospectively collected from the archive of the Department of Pathology of the Maastricht University Medical Center. Written informed consent was obtained from all study participants and the process adhered to local ethics guidelines.

Genotyping

To identify carriers of the *NDRG4* wild-type and deletion allele, purified DNA was examined by PCR. Genomic DNA was purified with the Gentra Puregene Mouse tail kit (Qiagen) according to manufacturer's instructions. For the detection of each allele, a PCR was performed with the *NDRG4* primer mix listed in Supplemental Table S5.1. The PCR reaction mix contained 100ng genomic DNA, 10 µl REDExtract-N-Amp PCR Reaction Mix (REDExtract-N-Amp™ Tissue PCR Kit, Sigma-Aldrich) and 0.8 µl primer mix (10 µM) in a final volume of 20 µl. The PCR was performed using the Biorad T100™ thermal cycler (Biorad) with the following conditions: 1) initial incubation: 94°C for 3 min, 2) 35 cycles: 94°C for 30 sec – 60°C (annealing temperature) for 30 secs – 72°C for 1 min, and 3) final elongation: 72°C for 10 min. PCR products were detected in a 1.5% (w/v) agarose gel in 0.5 × Tris-borate-EDTA (TBE) buffer. The *NDRG4* wild-type locus is identified by a 203 and 957bp fragment, whereas a 320bp band identifies the deletion allele.

Protein isolation and western blotting

Tissues (brain, heart and colon) from *NDRG4*$^{+/+}$, *NDRG4*$^{+/-}$ and *NDRG4*$^{-/-}$ mice were homogenized in RIPA buffer (Pierce technology) containing protease inhibitors (1 "complete" pill/50 ml, Roche), and resolved in SDS-gel electrophoresis. Protein transfers were probed overnight (4°C) with mouse anti-human NDRG4, clone 2G3 (1:500, H00065009-M01, Abnova), rabbit anti-human NDRG4 (1:1000, #9039, Cell Signaling) and β-actin (1:200 000, Sigma-Aldrich). Bound antibodies were visualized by an HRP-linked secondary anti-rabbit or anti-mouse antibody (1 hour, RT, Cell Signaling) and chemiluminescence (ECL, Pierce Biotechnology).

Immunohistochemistry

Three µm thick paraffin sections of *NDRG4*$^{+/+}$ and *NDRG4*$^{-/-}$ mouse tissues (brain, stomach, small intestine, colon, pancreas, heart, trachea, esophagus, spleen, liver, lungs, kidneys, adrenal glands, thymus, lymph nodes, skin, thyroid, salivary gland, gallbladder, bladder, uterus, testis, prostate, spinal cord) and human colon were deparaffinized in xylene and rehydrated in graded alcohols. To quench endogenous peroxidase activity, the slides were incubated with 0.3% hydrogen peroxide in methanol for 20 minutes. Antigen retrieval was performed by boiling the sections in Tris-EDTA buffer (Klinipath pH 8.0) or Dako target retrieval solution (Dako, pH 6.0), followed by blocking nonspecific antibody binding with PBS containing 20% fetal bovine serum and 0.1% Tween. Sections where incubated overnight at 4°C with the primary antibodies mouse anti-human NDRG4-clone 2G3 (1:500, Abnova) or rabbit anti-human NDRG4 antibody (Mouse brain and other organs 1:500 and 1:250, resp.; Human colon, 1:25; Cell signaling) diluted in PBS/0.5%BSA/0.1%Tween. After incubation with the biotinylated anti-mouse secondary antibody (1:250, RPN1001v1, Amersham Biosciences) followed by the Avidin-Biotin Complex detection method (ABC, 1:500, Dako Cytomation) or incubation with horseradish peroxidase-conjugated anti-rabbit IgGs (poly-HRP, Immunologic, Duiven, the Netherlands), bound antibodies were visualized using 3,3'-diaminobenzidine (DAB, Dako) as a chromogen (brown precipitate). Slides were counterstained with hematoxylin, dehydrated and mounted. In addition, to diminish recognition of endogenous mouse immunoglobins by the mouse primary antibody slides were subsequently stained with the mouse anti-human NDRG4 antibody, clone 2G3 (1:500, H00065009-M01, Abnova) using the Vector® M.O.M.™. Immunodetection Kit according to manufacturer's instructions (Vector Labs). Negative controls for primary antibodies were monitored by staining of *NDRG4*$^{-/-}$ mice tissues and omission of primary antibodies from *NDRG4*$^{+/+}$ slides. Images were acquired at RT using a Nikon DMX1200 digital camera and the ACT-1 v2.62 software from Nikon Corporation.

Immunofluorescence

Immunofluorescent stainings were performed on brain slices and whole-mount gut tissue as previously described.[18] Briefly, ileum and colon of *NDRG4*$^{+/+}$ and *NDRG4*$^{-/-}$ mice were collected, opened along the mesenteric border, stretched and pinned flat with insect pins (0.2 mm, Agar Scientific) in a sylgard lined dissection dish (Sylgard 184 Elastomer, Down Corning) with Krebs solution bubbled with 95% O_2 to 5% CO_2. The mucosal and submucosal layers were removed, tissues fixed for 30 minutes (4°C) in paraformaldehyde (4% in PBS) and rinsed in PBS. To visualize the myenteric and submucosal plexus the circular or longitudinal muscle layer, respectively, were peeled. Brain and gut tissues were permeabilized in 0.5% tritonX-100 and incubated in blocking solution (4% goat/donkey serum), followed by an overnight incubation (4°C) with the primary antibodies diluted in blocking solution: rabbit anti-NDRG4 (1:500, Cell signaling), chicken anti-GFAP (1:5000, Abcam), mouse anti-HuC/HuD (1:500, Invitrogen Life Technologies), sheep-anti neuronal NO-synthase (1:400, Santa Cruz Biotechnologies), rabbit anti-Calretinin (1:2000, Chemicon) and mouse anti-human NeuN (1:100) (Chemicon). After rinsing, tissues were incubated for 2 hours with the fluorescently labeled secondary antibodies: anti-human Alexa594, anti- mouse Alexa594, anti-rabbit Alexa488, anti-sheep Alexa488 (all 1:1000; Molecular probes, Invitrogen), anti-rabbit AMCA or anti-chicken AMCA (both 1:250; Jackson Immuno Research Labs). Samples were mounted with Vectashield™ mounting medium (Vector Labs). Preparations were imaged with a Zeiss LSM780 confocal microscope (Cell imaging Core, KU Leuven).

In situ mRNA hybridization

In situ mRNA hybridization was performed on formalin-fixed, paraffin-embedded brain tissue and intestinal Swiss rolls of *NDRG4*$^{+/+}$ and *NDRG4*$^{-/-}$ mice using digoxigenin (DIG) labeled riboprobes (DIG RNA Labeling Kit, SP6/T7, Roche) for NDRG4 nucleotide positions 1269-1777 (NDRG4-A) and 1811-2343 (NDRG4-B) of mouse NDRG4 (NM_145602, Supplemental Table S5.2).[8] Four µm sections were dried upright, deparaffinized, rehydrated and washed. Tissues were permeabilized with 0.1% pepsin in 0.2N HCl for 5 min (37°C) and post-fixed in 4% paraformaldehyde (4°C). Fixed sections were then treated twice with 100mM glycine in 1xPBS and pre-hybridized (37°C) for 45 min in pre-hybridization buffer (2xSCC, 50% (v/v) deionized formamide, 40% (v/v) DEPC-treated H_2O). *NDRG4* mRNA was detected overnight at 37°C with the pre-heated hybridization buffer (2x SCC, 50% (v/v) formamide, 10% dextran sulphate, 10mM DTT,7,5% DEPC-H_2O, 1x Denhardts solution, 1 mg/ml yeast tRNA, 1mg/ml denatured and sheared salmon sperm DNA) containing a mixture of both antisense or sense probes (1µg/ml). After a serial wash in descending concentrations (2x, 1x and 0.1x) of SCC buffer, tissue RNA was digested by RNase A (Roche), followed by a serial wash in ascending SCC concentrations (1x and 2x) and buffer 1 (100 mM Tris-HCl, 150

mM NaCl). Sections were then blocked for two hours (0.1% Triton X-100, 2% normal sheep serum in buffer 1) and incubated with 1:100 anti-DIG alkaline phosphatase Fab fragments (Roche) for one hour at 37°C. Following washes in buffer 1 and 2 (100mM Tris-HCl, 100 mM NaCl, 50mM $MgCl_2$), the chromogenic substrates nitro-blue tetrazolium chloride and 5-bromo-4chloro-3indolyl phosphate (Roche), dissolved in buffer 2 containing 5 mM levamisole, were used to detect hybridized alkaline phosphatase activity. Once the signal reached optimal intensity, the color reaction was stopped (10 mM Tris-HCl, 1 mM EDTA), sections were counterstained with nuclear fast red, clarified in tap water and mounted in imsol (1:3). Tissue sections of *NDRG4*$^{-/-}$ and *NDRG4*$^{+/+}$ mice immersed in antisense or sense riboprobes, resp. served as negative controls. Images were acquired as described for the immunohistohemistry images.

Results

The cell signaling antibody is able to specifically target NDRG4

We used *NDRG4* wild-type (*NDRG4*$^{+/+}$), heterozygous (*NDRG4*$^{+/-}$) and knockout (*NDRG4*$^{-/-}$) mice (which were characterized by genotyping analysis (Supplemental Figure S5.1)) to investigate the specificity of two commercially available antibodies, i.e. the monoclonal mouse anti-human NDRG4 antibody (clone 2G3, Abnova) and the polyclonal rabbit anti-human NDRG4 antibody (#9039, Cell Signaling).[16] Using Western Blot analysis we observed, as previously described[8,10,17], that NDRG4 is highly expressed in mouse brain and heart. Application of both antibodies showed that brain tissue of *NDRG4*$^{+/+}$ mice ubiquitously expressed the three well-known NDRG4 isoforms, NDRG4H (41 kDa), NDRG4B^{var} (39 kDa) and NDRG4B (37 kDa), and that *NDRG4*$^{+/+}$ hearts displayed a high expression of NDRG4H (Figure 5.1A-B, lane 1-4). Only the Cell Signaling antibody also detected NDRG4B^{var} and a relatively low level of NDRG4B within wild-type heart tissue (Figure 5.1A-B, lane 4). Interestingly, both antibodies identified NDRG4H in the proximal but not in the distal colon, whereas only the Cell Signaling antibody showed expression of NDRG4B^{var} in the proximal and distal colon of *NDRG4*$^{+/+}$ mice (Figure 5.1A-B, lane 7, 10). Likewise, brain, heart and colon tissue of *NDRG4*$^{+/-}$ mice displayed a similar pattern, except for slightly reduced expression levels compared to *NDRG4*$^{+/+}$ mice (Figure 5.1A-B, lane 2, 5, 8, 11). Successful ablation of the *NDRG4* gene was confirmed by the absence of all NDRG4 isoforms, (indicated by the lack of bands/signals), in all tissues of *NDRG4*$^{-/-}$ mice after application of the Cell Signaling antibody. However, the specificity of the Abnova antibody could not be confirmed as the signal was reduced, but still detectable in all tissues of *NDRG4*$^{-/-}$ mice (Figure 5.1A-B, lane 3, 6, 9, 12). Furthermore, the Abnova antibody detected various non-specific bands of about 50-55 kDa in all tissues (Figure 5.1A, all lanes). In addition, using immunohistochemistry with the Abnova antibody, *NDRG4*$^{-/-}$ mice exhibited non-specific staining in brain neuronal cells, in cardiomyocytes and in intestinal epithelial cells

similar to the pattern observed in *NDRG4*$^{+/+}$ tissue (Figure 5.1C). The Cell Signaling antibody, on the other hand, was able to accurately detect NDRG4. Neurons of the brain, heart and intestines of *NDRG4*$^{+/+}$ mice showed a strong reactivity to this antibody, which was completely absent in the *NDRG4*$^{-/-}$ mice (Figure 5.1D). In addition, *in situ* mRNA hybridization on different tissues confirmed the NDRG4 expression pattern observed by the Cell Signaling antibody (data not shown). In conclusion, the Cell Signaling antibody meets the criteria to specifically target NDRG4 and was therefore used throughout this study to investigate the whole-body expression of NDRG4.

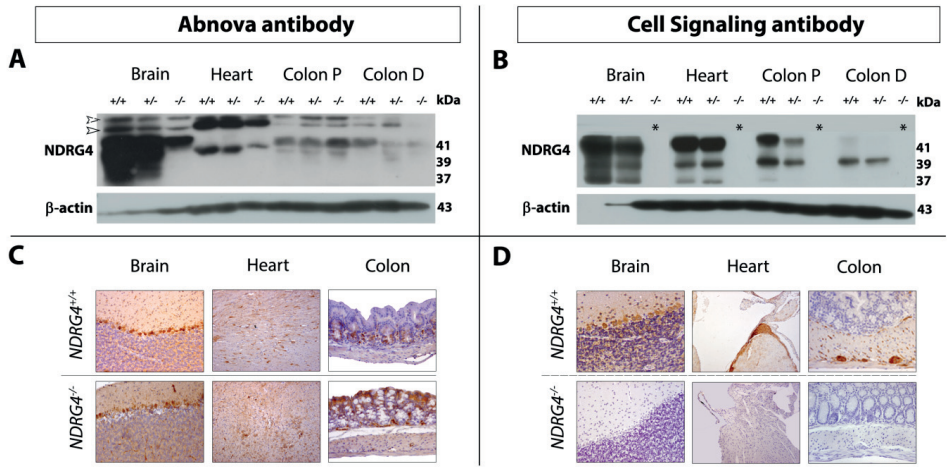

Figure 5.1 The Cell signaling, but not the Abnova antibody specifically targets NDRG4 according the specificity criteria.

(**A-B**) Expression pattern of NDRG4 in mouse brain, heart and colon (P, proximal; D, distal). Equal protein amounts of organ homogenates of one-year old *NDRG4*$^{+/+}$, *NDRG4*$^{+/-}$ and *NDRG4*$^{-/-}$ mice were subjected to Western Blotting using the Abnova (A) and Cell signaling antibody (B). Both antibodies recognize NDRG4B (37kDa), NDRG4B^{var} (39kDa) and NDRG4H (41kDa) in *NDRG4*$^{+/+}$ and *NDRG4*$^{+/-}$ tissues. β-actin is used as loading control. * Indicates successful detection of *NDRG4* ablation and open triangles indicate non-specific bands (50-55kDa). (**C-D**) Brain (200x), heart (100x) and colon (200x) sections from one-year old *NDRG4*$^{+/+}$ and *NDRG4*$^{-/-}$ mice were subjected to immunohistochemical analysis with the Abnova (C) and Cell Signaling antibody (D). Positive signals (brown) detected by the Abnova antibody are not affected by absence of *NDRG4* and differ from the highly selective staining pattern produced by the Cell signaling antibody.

NDRG4 is expressed by neuronal cells in the central and peripheral nervous system

Consistent with literature on the central nervous system, gray matter structures in the cerebrum and cerebellum of *NDRG4*$^{+/+}$ mice showed diffuse, predominantly cytoplasmic NDRG4 expression (Figure 5.2A). More specifically, in the hippocampal

cortex, NDRG4 was expressed in the pyramidal neurons and the granule cells of the dentate gyrus, which are characterized by NeuN reactivity (Figure 5.2A, NDRG4 in green and NeuN in red). Likewise, various NeuN positive neurons in the thalamus and medulla oblongata express NDRG4 (Figure 5.2A, thalamus). Furthermore, weak NDRG4 expression was detected in the neuropil, a meshwork of merely unmyelinated axons, dendrites and glial processes associated with the cerebral cortex (data not shown). In the cerebellar cortex, the three layers, i.e. molecular-, Purkinje- and granular layer, showed NDRG4 expression. Most striking was the ubiquitous expression of NDRG4 in Purkinje cells (which are negative for NeuN labeling) and their processes throughout the molecular and granular layer. In addition, the basket cells of the molecular layer were strongly positive for NDRG4 (Figure 5.2A, Purkinje layer). In contrast, the white matter of the cerebrum and corpus callosum, as well as the ependymal cells and meninges were negative for NDRG4 (data not shown). In representative sections of *NDRG4*$^{-/-}$ brain, absence of NDRG4 expression was confirmed (Figure 5.2A). Similar to the brain, the ventral and dorsal root ganglia and neural cell bodies in the gray matter of the spinal cord, showed strong cytoplasmic NDRG4 reactivity, while the white matter, ependymal cells and meninges were negative for NDRG4 (Figure 5.2A).

In the peripheral nervous system (PNS), the cytoplasm of ganglia and neuronal cells, together with the nerve fibers to, from and in every organ of the body (stomach, small intestine, colon, pancreas, heart, trachea, esophagus, spleen, liver, lungs, kidneys, adrenal gland, thymus, lymph nodes, skin, thyroid, salivary gland, gallbladder, bladder, uterus, testes, prostate) strongly expressed NDRG4 (Figure 5.2B, heart and lung). In contrast to previously published data[11,17,19]. NDRG4 expression was not detected within cardiomyocytes, but specifically localized in cardiac Purkinje fibers. Finally, the perivascular nerve fibers in the tunica adventitia of arteries and veins, also known as the nervi vascularis, showed weak but consistent expression of NDRG4 (data not shown). In summary, NDRG4 is predominantly expressed within neurons of the central and peripheral nervous system.

NDRG4 is specifically expressed within the enteric nervous system of the mouse and human gastro-intestinal tract

Characterization of NDRG4 expression in the gastro-intestinal (GI) tract of *NDRG4*$^{+/+}$ and *NDRG4*$^{-/-}$ mice using immunohistochemistry, revealed that NDRG4 is specifically expressed within the nervous system of the gut, i.e. the enteric nervous system (ENS; Figure 5.3). More precisely, NDRG4 is expressed in the cytoplasm of cell bodies inside ganglia of the myenteric (Auerbach's) plexus, located between the outer longitudinal and inner circular muscle layer along the entire GI-tract (Figure 5.3A, black arrow), and within the ganglia of the submucosal (Meissners) plexus (Figure 5.3A, red arrow).

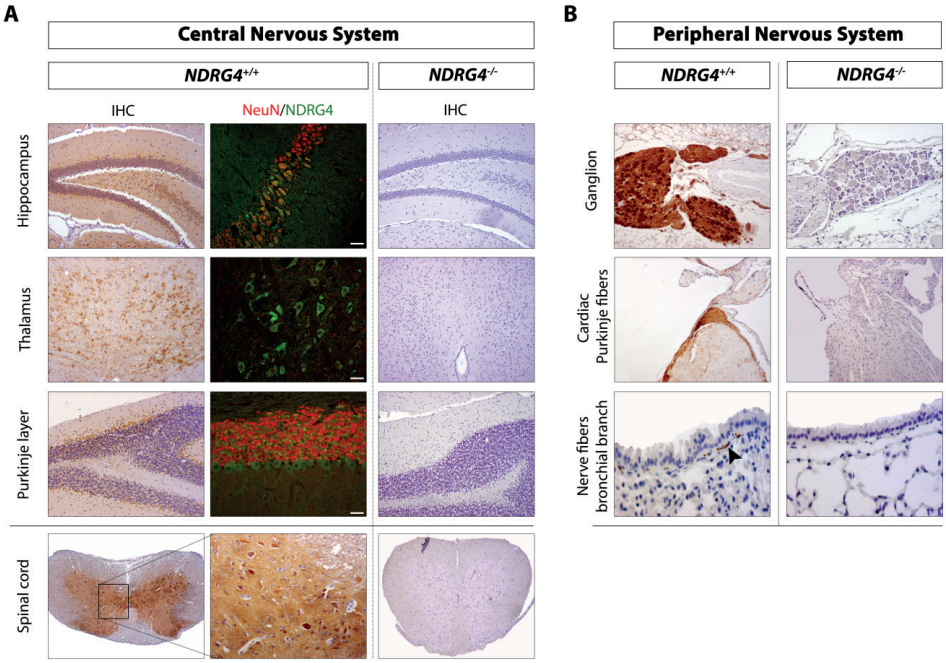

Figure 5.2 **NDRG4 is specifically expressed within the central and peripheral nervous system.** (**A**) Immunohistochemical (100x magnification) and immunofluorescent analysis (scale bar, 20µm) of several areas of the one-year old mouse brain. In various regions of the cerebrum: e.g. hippocampus and thalamus, NDRG4 (green) colocalizes with the neuronal marker NeuN (red). Cerebral Purkinje cells (negative for NeuN) are strongly positive for NDRG4. The gray matter of the spinal cord shows NDRG4 reactivity within neural cell bodies (brown, 40x and 200x magnification). NDRG4 is absent in all CNS structures in *NDRG4*$^{-/-}$ mice. (**B**) Representative images of peripheral tissues display NDRG4 expression in ganglions, neurons, Purkinje fibers (all 100x magnification), and in nerves to, from and within every organ of the body (400x magnification).

The nerve fiber bundles connecting the ganglia in both plexus also showed NDRG4 expression. Furthermore, the tiny nerves innervating the outer and inner muscularis externa, muscularis mucosae and those projecting into the mucosa also showed strong and weaker NDRG4 positivity, respectively. (Figure 5.3A). Representative immunohistochemical images (Figure 5.3A) show that the intestinal tract of *NDRG4*$^{-/-}$ mice still harbors all the above described ENS structures, i.e. ganglia of the myenteric and submucosal plexus and interconnecting nerve fibers, but does not display NDRG4 reactivity, confirming the absence of NDRG4 expression. Furthermore, *in situ* mRNA hybridization confirmed that NDRG4 is specifically expressed within the ENS of *NDRG4*$^{+/+}$ mice (Figure 5.3A).

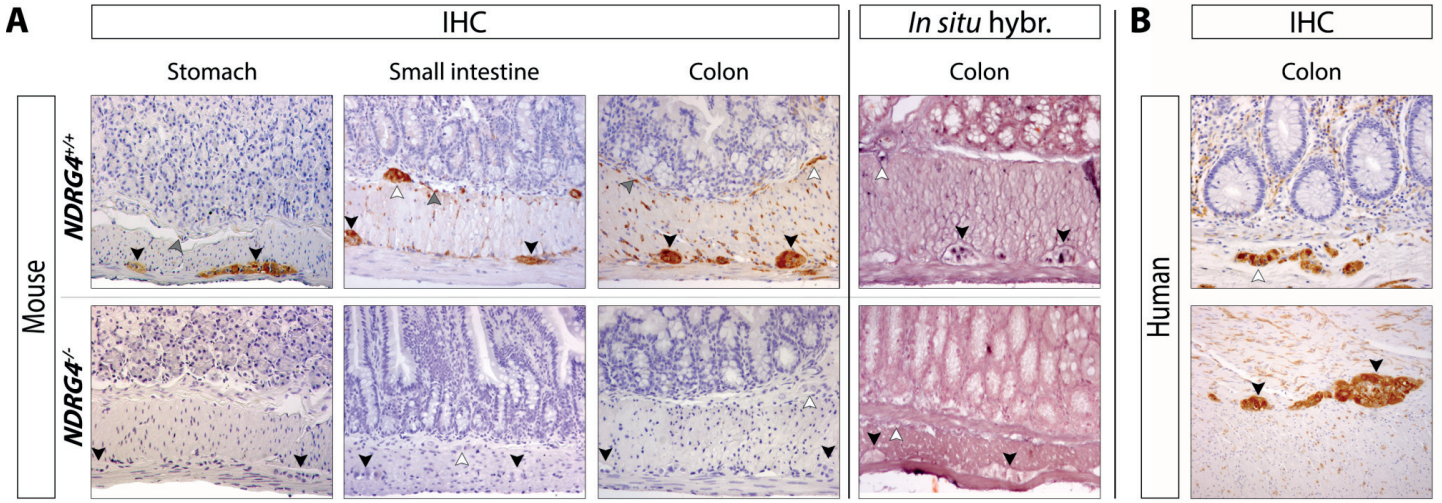

Figure 5.3 **NDRG4 is localized within the enteric nervous system of mouse and man**.
(**A**) NDRG4 protein and mRNA expression in *NDRG4+/+* and *NDRG4-/-* stomach, small and large intestine. Representative immunohistochemistry and in *in situ* mRNA hybridization reveals the presence of NDRG4 at protein and RNA level, resp. within ganglia of the myenteric (black arrow) and submucosal (red arrow) plexus, but also in interconnecting nerve fibers throughout the muscle layer and within the fiber bundles connecting the ganglia (200x magnification). Successful ablation of *NDRG4* is confirmed at RNA and protein level. (**B**) The NDRG4 expression pattern in human colonic specimens (100x magnification) is analogous as in mice.

Similar as described in mouse, human intestinal specimens displayed NDRG4 immunoreactivity in the ganglia of both plexus, in nerve fibers connecting these ganglia and within the fibers innervating both muscle layers of the muscularis externa and muscularis mucosae, as shown in Figure 5.3B. Compared to mouse, the mucosal projecting nerve fibers showed more robust NDRG4 immunoreactivity. Together these data indicate that NDRG4 is specifically expressed within the ENS of both mice and men.

NDRG4 is expressed within enteric neuronal cells

Whole mount preparations of both plexus of the *NDRG4*$^{+/+}$ and *NDRG4*$^{-/-}$ small and large intestine were used to investigate which cells of the ENS express NDRG4. Since similar results were observed in all preparations, only data from the colonic myenteric plexus are shown (Figure 5.4). Immunofluorescent triple-labeling revealed that within the gut of *NDRG4*$^{+/+}$ mice, NDRG4 is exclusively expressed within the cytoplasm of enteric neurons, as NDRG4 positive cells were always labeled for the pan-neuronal marker HuC/D, but never co-expressed the glial marker GFAP (Figure 5.4A, green, red and blue, resp.). Even though NDRG4 localized in enteric neurons throughout all areas of the ENS, not all enteric neurons (HuC/D positive) expressed NDRG4 (Figure 5.4A). Therefore, two main neuronal subsets, positive for the neuronal markers nitric oxide synthase (nNOS; inhibitory neurons) and calretinin (excitatory neurons), were further investigated (Figure 5.4B-C, green and blue resp). *NDRG4*$^{+/+}$ and *NDRG4*$^{-/-}$ colon have a similar neuronal population as indicated by analogous HuC/D positivity (Figure 5.4B, red)) and calretinin and nNOS staining (Figure 5.4B, blue and green resp.). Further analysis revealed that the NDRG4 (purple) and nNOS (green) positive population were, despite their comparable proportion (30% and 40% of the entire neuronal population resp.), almost completely different from each other (Figure 5.4C). More precisely, the expression of NDRG4 is inversely related to nNOS positivity: only 10% of the nNOS population was positive for NDRG4 and conversely, only 3% of the NDRG4 positive enteric neurons also expressed nNOS. In conclusion, NDRG4 is solely expressed within enteric neurons, but not in all enteric neurons since not every HuC/D positive cell expresses NDRG4.

Expression of NDRG4 outside the nervous system

Besides the very specific expression pattern of NDRG4 within central, peripheral and enteric nervous system structures, we also observed cytoplasmic NDRG4 expression outside the nervous system (Figure 5.5). In the adrenal gland, chromaffin cells (neuroendocrine cells) of the adrenal medulla strongly expressed NDRG4. In addition, the cells lining the collecting ducts of Bellini in the renal papilla also showed NDRG4 expression. The adrenal and renal cortex and renal medulla on the other hand were negative for NDRG4.

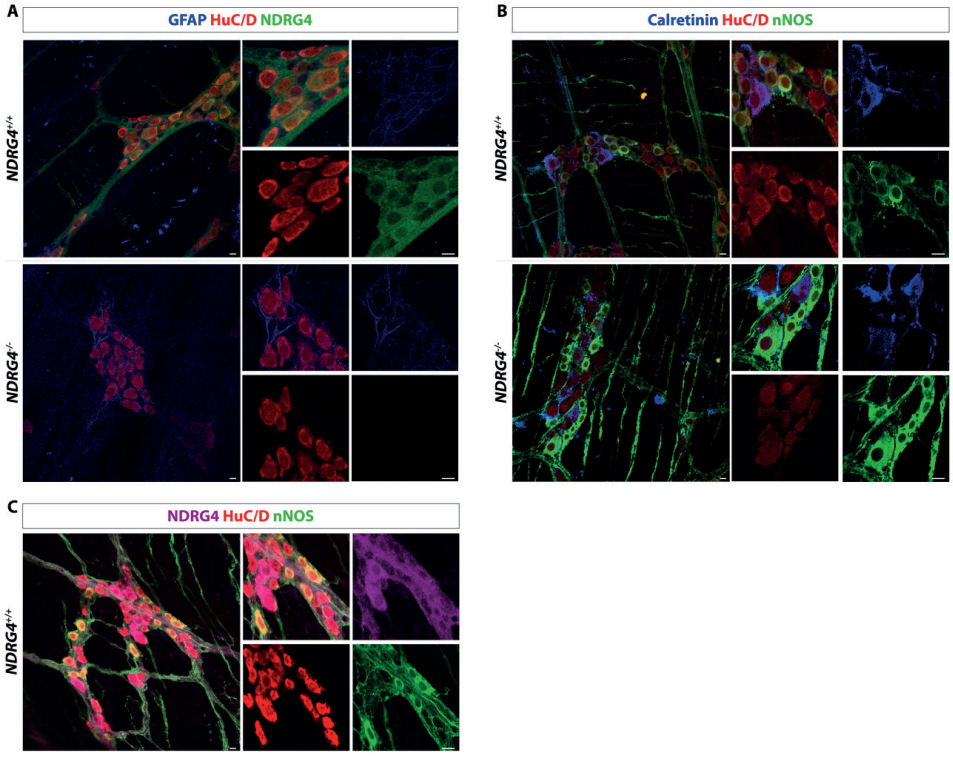

Figure 5.4 **NDRG4 is solely expressed within enteric neurons.**
Representative images of fluorescently labeled whole-mount preparations of colonic myenteric plexus of two month old *NDRG4*$^{+/+}$ and *NDRG4*$^{-/-}$ mice (scale bars, 5µm). (**A**) Immunofluorescent double labeling reveals that NDRG4 always localized with the pan-neuronal marker HuC/D (green and red, resp.) but never with the glial marker GFAP (green and blue, resp.). *NDRG4*$^{-/-}$ mice still have enteric neurons and glia cells (red and blue, resp.), despite the absence of NDRG4. (**B**) The plexus of *NDRG4*$^{+/+}$ and *NDRG4*$^{-/-}$ colon have a similar neuronal population as indicated by analogous HuC/D, calretinin and nNOS positivity (red, blue and green). (**C**) NDRG4 positive cells (purple) are always reactive to HuC/D (red), but almost never positive for nNOS (green), suggesting a negative correlation between NDRG4 and nNOS expression.

In the testis, NDRG4 was detected in several, but not all Sertoli cells, while Leydig cells exhibited no NDRG4 positivity. Finally, as described above, NDRG4 expression in the GI-tract was restricted to enteric neurons, except for NDRG4 reactivity within the chief cell population located in the basal regions of the fundic glands in the glandular stomach. All other analyzed organs (small intestine, colon, pancreas, heart, trachea, esophagus, spleen, liver, lungs, thymus, lymph nodes, skin, thyroid, salivary gland, gallbladder, bladder, uterus, prostate) were negative for NDRG4, except for the above mentioned NDRG4 expression in neuronal tissues.

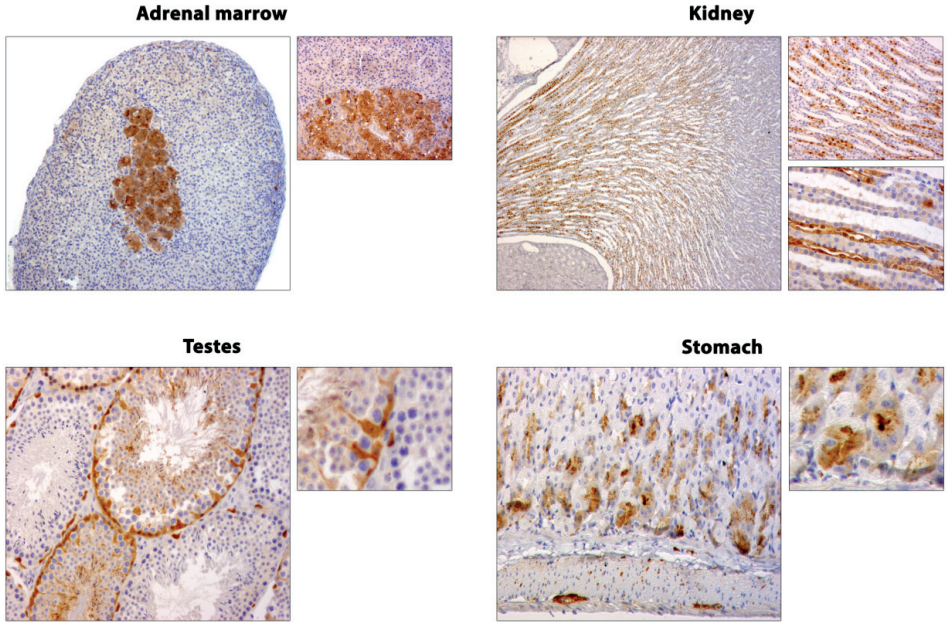

Figure 5.5 **NDRG4 is expressed within (neuro)endocrine related tissues outside the nervous system.** Representative histological images of the adrenal gland, kidney, testis and stomach of one-year *NDRG4*$^{+/+}$mice (100x magnification). The higher magnification (200x) image of the adrenal medulla clearly reveals strong NDRG4 immunoreactivity within chromaffin cells. Higher magnification images of the kidney (200x and 400x) display NDRG4 expression within cells lining the collecting ducts of Bellini. Several, but not all Sertoli cells in the testis and the chief cell population in the basal fundic gland in the glandular stomach show NDRG4 positivity (200x).

Discussion

With this study, we confirmed the neuronal specific expression of NDRG4 within the CNS and observed that NDRG4 positive cells always co-expressed the neuronal marker NeuN, but never localized with the glial cell marker GFAP. The strong expression patterns of NDRG4 within numerous regions of the brain, including hippocampus, thalamus and Purkinje cells, but also within the spinal cord, suggest an important role in development and diseases of the CNS. NDRG4 might protect against neurological deficits as NDRG4 has been shown to be necessary for normal hindbrain development, proper neuronal functioning and myelination of axons.[12,20] Even though the molecular mechanisms by which NDRG4 provides neuroprotection are unknown, NDRG4-mediated regulation of BDNF secretion is very likely.[8] It has been described that NDRG4 is essential in retaining levels of the neuroprotective factor BDNF[8] and that BDNF secretion is dependent on the influx of Calcium (Ca^{2+}).[21] In addition NDRG4 is necessary

for clustering of Ca^{2+} channels along axons[13], and SMAP8, the rat ortholog of NDRG4, is able to increase intracellular Ca^{2+} levels within hypothalamic neurons.[22] Consistently, a destabilization of Ca^{2+} homeostasis seems central in the pathogenesis of Alzheimer disease[23] and the Alzheimer diseased brain is characterized by reduced levels of both NDRG4 and BDNF.[6,8,24] All together these data indicate that NDRG4 is important within the CNS, but the mechanisms of action are still subject for further investigation.

Next, we characterized NDRG4 expression outside the CNS. In the heart, NDRG4H was the most abundant isoform detected by Western blotting. In addition, we also detected lower levels of NDRG4B and NDRG4B^{var}, which might be attributed to the high sensitivity of the chemiluminescent detection method on film of our Western blotting analysis. Using immunohistochemistry, we only detected NDRG4 within specialized subendocardial myocytes, the so-called Purkinje fibers, but not within cardiomyocytes. The Purkinje fibers are part of the pulse conducting system which regulates heart contractility and heart rate. Consistently, hearts of *NDRG4* deficient zebrafish are characterized by a reduced circulation and a slower heart rate with weaker contractility.[11] Similarly, single nucleotide polymorphisms in locus 16q21 near *NDRG4,* predispose to sudden cardiac death, as these variations have been associated with prolongation of the myocardial repolarization time (QT interval).[15]

Every other organ predominantly displayed NDRG4 immunoreactivity in their associated ganglia, nerve fibers and nerve bundles as part of the PNS. In addition to expression in the PNS, NDRG4 is expressed within specific cell types of the testis, adrenal/renal medulla and stomach. Within the testis, NDRG4 localized to Sertoli cells (also known as sustentacular cells), which serve an endocrine role by the production of various hormones (e.g. inhibin) and expression of a set of hormone receptors (e.g. androgen receptor). Cells of the collecting ducts (of Bellini) in the renal papilla were also NDRG4 positive. These cells resorb various salts and water during urine production which is mediated by endocrine hormones like aldosterone and vasopressin. NDRG4 positive chief cells of the glandular stomach are part of the endocrine system as they secrete the “satiety hormone” leptin. Finally, the adrenal medulla, a structure similar to a sympathetic ganglion without post-ganglionic processes, exposed NDRG4 positive chromaffin cells. These cells are neuro-crest derived, have the potential to develop into postganglionic sympathetic neurons and are able to synthetize catecholamine neurotransmitters. Together these results suggest a role for NDRG4 within the (neuro-) endocrine related system, but it is still unclear in what pathway NDRG4 might be involved.

Additionally, we explored NDRG4 expression within the GI-tract and revealed the presence of the NDRG4B^{var} and NDRG4H isoform in mouse colon and the specific expression of NDRG4 within the enteric nervous system. The ENS, the so-called brain of the gut, is an interconnected network of enteric neurons and glia cells, predominantly clustered in ganglia of the submucosal (Meissner’s) and myenteric (Auerbach’s) plexus, along the entire GI-tract.[25] The interplay between neurons and glial cells ensures proper

functioning of the ENS. Here we describe for the first time that NDRG4 expression in the gut is restricted to enteric neurons as NDRG4 always co-localized with the pan-neuronal marker HuC/D, but never with the glial cell marker GFAP. NDRG4 has already been identified by DNA microarray analysis and *in situ* hybridization as one of the significantly down-regulated genes in the aganglionic bowel of mice (*Ret*$^{k-/k-}$) and Hirschsprung patients.[26,27] Hirschsprung disease is characterized by the absence of the ENS, mainly in the distal colon, leading to severe, sometimes life-threatening constipation due to impaired gut homeostasis. In addition to enteric neuropathies, the importance of the ENS in the GI-tract is further highlighted by various digestive diseases and clinical symptoms (e.g. inflammatory bowel disease (IBD), Crohn's disease) that arise when alterations in the ENS occur. It has been described that the density of enteric neurons/nerves determines the severity of colitis, i.e. transgenic mice characterized by a greater than normal number of enteric neurons (NSE-noggin) suffer from more severe intestinal inflammation compared to mice with fewer than normal neurons (*Hand2*$^{+/-}$).[28] Neurons synthesize, like endocrine cells, various neurotrophic factors and neurotransmitters, which are packaged in vesicles and released in exocytotic processes. The process of exocytosis allows a cell to communicate with neighboring cells, i.e. other neurons, glia, muscle, endothelial, immune and epithelial cells.[25] Thus, apart from providing intrinsic innervation to the muscular layers, the ENS is also important to preserve intestinal homeostasis by controlling cells within the intestinal epithelium and communicating with the endocrine and immune system. NDRG4 has been implicated in the route of vesicular transport. The interaction of NDRG4 with Blood vessel epicardial substance (Bves) has been shown to regulate docking of VAMP-3 (SNARE-protein) positive vesicles to the cell surface and the subsequent cargo delivery (e.g. fibronectin).[19] Similarly, NDRG4 controls vesicle membrane fusion during exocytosis as *NDRG4* knockdown has been associated with a sharp reduction in the level of another vesicle-SNARE protein, SNAP25.[13] The SNARE proteins, including SNAP25, are complexes that are essential in promoting subcellular trafficking, vesicular fusion and subsequent exocytotic release of neurotransmitters.[29] Therefore NDRG4, like NDRG1 and NDRG2, might modulate subcellular vesicle trafficking and exocytotic release of neuromediators,[30,31] thereby regulating neurotransmission and the subsequent targeting of the epithelial layer.

We and others, previously observed decreased NDRG4 expression in the intestinal epithelium in the transition towards CRC.[1,32-34] In these studies, the commercially available monoclonal mouse anti-human NDRG4 from Abnova (clone 2G3, H00065009-M01) was used, which is, to our knowledge, never validated to specifically target NDRG4. In this study we compared two commercially available antibodies: the above-mentioned monoclonal mouse anti-human NDRG4 from Abnova (clone 2G3, H00065009-M01[1,17,34]) and the polyclonal rabbit anti-human NDRG4 from Cell Signaling (#9039[35,36]) according to the specificity criteria.[16] Western blotting and immunohistochemical analysis, using tissues of *NDRG4*$^{+/+}$ and *NDRG4*$^{-/-}$ mice, revealed

that successful ablation of *NDRG4* could only be confirmed using the Cell Signaling antibody. In addition, the expression pattern observed by application of the Cell signaling antibody was confirmed by mRNA *in situ* hybridization. Given the high sequence similarity (52-65%) between NDRG4 and its protein family members, it is likely that the Abnova antibody recognizes one of the other NDRG proteins.

In summary, we observed that NDRG4 is specifically expressed within the neurons of the CNS, PNS and ENS and is not expressed in colonic epithelial cells. It remains to be elucidated how this expression pattern is linked with its functional role(s) and the CRC biomarker potential in gut.

References

1. Melotte V, et al. N-Myc downstream-regulated gene 4 (NDRG4): a candidate tumor suppressor gene and potential biomarker for colorectal cancer. J Natl Cancer Inst 2009:101(13):916-927.
2. Ahlquist DA, et al. The stool DNA test is more accurate than the plasma septin 9 test in detecting colorectal neoplasia. Clin Gastroenterol Hepatol 2012;10(3):272-277 e1.
3. Ahlquist DA, et al. Next-generation stool DNA test accurately detects colorectal cancer and large adenomas. Gastroenterology 2012;142(2):248-256; quiz e256.
4. Imperiale TF, et al. Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med 2014; 370(14):1287-97.
5. Lidgard GP, et al. Clinical performance of an automated stool DNA assay for detection of colorectal neoplasia. Clin Gastroenterol Hepatol 2013;11(10):1313-1318.
6. Zhou RH, et al. Characterization of the human NDRG gene family: a newly identified member, NDRG4, is specifically expressed in brain and heart. Genomics 2001;73(1):86-97.
7. Qu X, et al. Characterization and expression of three novel differentiation-related genes belong to the human NDRG gene family. Mol Cell Biochem 2002;229(1-2):35-44.
8. Yamamoto H, et al. NDRG4 protein-deficient mice exhibit spatial learning deficits and vulnerabilities to cerebral ischemia. J Biol Chem 2011;286(29):26158-165.
9. Nakada N, et al. Molecular characterization of NDRG4/Bdm1 protein isoforms that are differentially regulated during rat brain development. Brain Res Dev Brain Res 2002;135(1-2):45-53.
10. Okuda T, Kokame K, Miyata T. Differential expression patterns of NDRG family proteins in the central nervous system. J Histochem Cytochem 2008;56(2):175-182.
11. Qu X, et al. Ndrg4 is required for normal myocyte proliferation during early cardiac development in zebrafish. Dev Biol 2008;317(2):486-496.
12. Ohki T, et al. Inhibition of neurite outgrowth by reduced level of NDRG4 protein in antisense transfected PC12 cells. Brain Res Dev Brain Res 2002;135(1-2):55-63.
13. Fontenas L, Chambraud B, Tawk M. Neuronal NDRG4 is essential for nodes of Ranvier organization and myelination in zebrafish. in European Meeting on Glial Cells in Health and Disease. 2015. Bilbao.
14. Dupays L, et al. Tbx2 misexpression impairs deployment of second heart field derived progenitor cells to the arterial pole of the embryonic heart. Dev Biol 2009;333(1):121-131.
15. Newton-Cheh C, et al. Common variants at ten loci influence QT interval duration in the QTGEN Study. Nat Genet 2009;41(4):399-406.
16. Pradidarcheep W, et al. Lack of specificity of commercially available antisera: better specifications needed. J Histochem Cytochem 2008;56(12):1099-1111.
17. Qu X, Li J, Baldwin HS. Postnatal lethality and abnormal development of foregut and spleen in Ndrg4 mutant mice. Biochem Biophys Res Commun 2016;470(3):613-619.
18. Herdewyn S, et al. Prevention of intestinal obstruction reveals progressive neurodegeneration in mutant TDP-43 (A315T) mice. Mol Neurodegener 2014;9:24.
19. Benesh EC, et al. Bves and NDRG4 regulate directional epicardial cell migration through autocrine extracellular matrix deposition. Mol Biol Cell 2013;24(22):3496-3510.
20. Fontenas L, Chambraud B., Tawk M. Neuronal NDRG4 is essential for nodes of Ranvier organization and myelination in zebrafish. in Poster abstract, XII European Meeting on Glial Cells in Health and Disease. 2015. Bilbao.
21. Kolarow R, Brigadski T, Lessmann V. Postsynaptic secretion of BDNF and NT-3 from hippocampal neurons depends on calcium calmodulin kinase II signaling and proceeds via delayed fusion pore opening. J Neurosci 2007;27(39):10350-10364.
22. Brailoiu GC, et al. Smooth muscle-associated protein 8: distribution and biological activity in the rat brain. J Neurosci Res 2007;85(8):1789-1796.
23. LaFerla FM, Calcium dyshomeostasis and intracellular signalling in Alzheimer's disease. Nat Rev Neurosci 2002;3(11):862-872.
24. O'Bryant SE, et al. Brain-derived neurotrophic factor levels in Alzheimer's disease. J Alzheimers Dis 2009;17(2):337-341.

25. Furness JB. The enteric nervous system and neurogastroenterology. Nat Rev Gastroenterol Hepatol 2012;9(5):286-294.
26. Heanue TA, Pachnis V. Expression profiling the developing mammalian enteric nervous system identifies marker and candidate Hirschsprung disease genes. Proc Natl Acad Sci U S A 2006;103(18): 6919-6924.
27. Vohra BP, et al. Differential gene expression and functional analysis implicate novel mechanisms in enteric nervous system precursor migration and neuritogenesis. Dev Biol 2006;298(1):259-271.
28. Margolis KG, et al. Enteric neuronal density contributes to the severity of intestinal inflammation. Gastroenterology 2011;141(2):588-598, 598 e1-2.
29. Tafoya LC, et al. Expression and function of SNAP-25 as a universal SNARE component in GABAergic neurons. J Neurosci 2006;26(30):7826-7838.
30. Choi SC, et al. NDRG2 is one of novel intrinsic factors for regulation of IL-10 production in human myeloid cell. Biochem Biophys Res Commun 2010;396(3):684-690.
31. Kachhap SK, et al. The N-Myc down regulated Gene1 (NDRG1) Is a Rab4a effector involved in vesicular recycling of E-cadherin. PLoS One 2007;2(9):e844.
32. Kim YJ, et al. NDRG2 suppresses cell proliferation through down-regulation of AP-1 activity in human colon carcinoma cells. Int J Cancer 2009;124(1):7-15.
33. Kovacevic Z, Richardson DR. The metastasis suppressor, Ndrg-1: a new ally in the fight against cancer. Carcinogenesis 2006;27(12):2355-2366.
34. Chu D, et al. NDRG4, a novel candidate tumor suppressor, is a predictor of overall survival of colorectal cancer patients. Oncotarget 2015;6(10):7584-7596.
35. Kolodziej MA, et al. NDRG2 and NDRG4 Expression Is Altered in Glioblastoma and Influences Survival in Patients with MGMT-methylated Tumors. Anticancer Res 2016;36(3):887-897.
36. Xing Y, et al. N-myc downstream-regulated gene 4, up-regulated by tumor necrosis factor-alpha and nuclear factor kappa B, aggravates cardiac ischemia/reperfusion injury by inhibiting reperfusion injury salvage kinase pathway. Basic Res Cardiol 2016;111(2):11.

Supplementary figure and tables

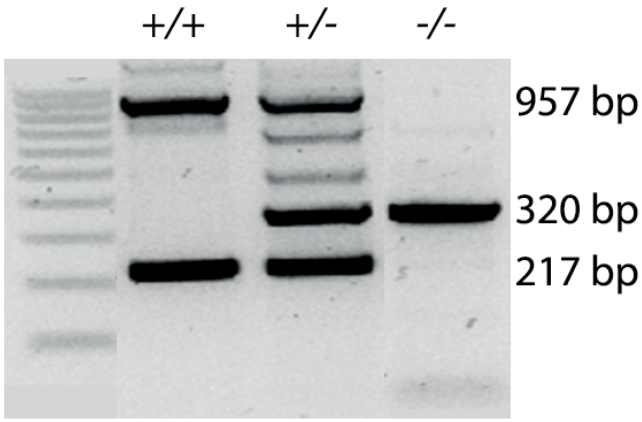

Figure S5.1 **Genotyping PCR analysis differentiates *NDRG4*$^{+/+}$, *NDRG4*$^{+/-}$ and *NDRG4*$^{-/-}$ mice.** *NDRG4*$^{+/+}$ mice show wild-type fragments of 203 and 975bp (**lane 1**), *NDRG4*$^{-/-}$ mice display a 320bp band for the deletion allele (**lane 3**) and *NDRG4*$^{+/-}$ mice show both the wild-type and deletion fragments (**lane 2**).

Table S5.1 **PCR primers for genotyping.**

	Target	Forward primer	Reverse primer	Annealing temp. (°C)
NDRG4 Primer mix	NDRG4 4HLOX1	TAGGCAGGGGCAGGTGGGTTTGT		
	NDRG4 4HLOX2		GGCGTCCTGATGTCATGTTCCTGT	60
	NDRG4 4H776		GCTCCCACTCCAATGCCAATC	

Table S5.2 **Probes for *in situ* mRNA hybridization.**

Target NDRG4 mouse	Forward primer	Reverse primer	Annealing temp. (°C)
NDRG4-A	TTATTTAAAAAGAAATGAGGGGATC	TTGCCTCAGGGTGGGACAA	58.5 (F) – 60.2 (R)
NDRG4-B	GTTAAAATGTTGATTGCTGTGTATGC	ACTCCAGAGCAGTCTAGAAATGGC	60.5 (F) – 60.9 (R)

Chapter 6

Spectrin repeat containing nuclear envelope 1 and *Forkhead box protein E1* are promising markers for the detection of colorectal cancer in blood

Veerle Melotte*, Joo Mi Yi*, Marjolein HFM Lentjes, Kim M Smits, Leander Van Neste, Hanneke EC Niessen, Kim AD Wouters, Joost Louwagie, Kornel E Schuebel, James G. Herman, Stephen B Baylin, Wim van Criekinge, Gerrit A. Meijer, Nita Ahuja*, Manon van Engeland*

Cancer Prev Res (Phila) 2015;8(2):157-164

**authors contributed equally to this work*

Abstract

Identifying biomarkers in body fluids may improve the non-invasive detection of colorectal cancer (CRC). Previously we identified *N-Myc downstream-regulated gene 4* (*NDRG4*) and *GATA binding protein 5* (*GATA5*) methylation as promising biomarkers for CRC in stool DNA. Here, we examined the utility of *NDRG4*, *GATA5*, and two additional markers (*Forkhead box protein E1* (*FOXE1*) and *spectrin repeat containing nuclear envelope 1* (*SYNE1*)) promoter methylation as biomarkers in plasma DNA. Quantitative methylation-specific PCR was performed on plasma DNA from 220 CRC patients and 684 non-cancer controls, divided in a training and test set. Receiver operating characteristic analysis was performed to measure the area under the curve of *GATA5*, *NDRG4*, *SYNE1* and *FOXE1* methylation. Functional assays were performed in *SYNE1* and *FOXE1* stably transfected cell lines. The sensitivity of *NDRG4*, *GATA5*, *FOXE1* and *SYNE1* methylation in all stages of CRC (154 cases, 444 controls) was 27% (95%CI=20%-34%), 18% (95%CI=12%-24%), 46% (95%CI=38%-54%) and 47% (95%CI=39%-55%), with a specificity of 95% (95%CI=93%-97%), 99% (95%CI=98%-100%), 93% (95%CI=91%-95%) and 96% (95%CI=94%-98%) respectively. Combining *SYNE1* and *FOXE1*, increased the sensitivity to 56% (95%CI=48%-64%), while the specificity decreased to 90% (95%CI=87%-93%) in the training set and to 58% sensitivity (95%CI=46%-70%) and 91% specificity (95%CI=80%-100%) in a test set (66 cases, 240 controls). *SYNE1* overexpression showed no major differences in cell proliferation, migration and invasion compared to controls. Overexpression of *FOXE1* significantly decreased the number of colonies in SW480 and HCT116 cell lines. Overall our data suggests that *SYNE1* and *FOXE1* are promising markers for CRC detection.

Introduction

Early detection of colorectal cancer (CRC) offers opportunity to cure CRC.[1] Colonoscopy is the gold standard for CRC screening, however this procedure is invasive, expensive, and not readily accessible or acceptable to the majority of age-eligible individuals.[2] Therefore, the search for non-invasive screening methods has intensified. Currently, fecal immunochemical test (FIT) is used as a non-invasive and relatively inexpensive CRC screening modality. In addition, potential biomarkers in stool DNA have been described. Despite the high sensitivity and specificity of some of these markers in stool DNA, we hypothesized that a blood-based test, not depending on stool sampling, has the potential for better patient compliance and is better suited for systems without programmatic screening.

Epigenetic silencing of tumor suppressor genes by aberrant promoter methylation frequently occurs in human cancers.[3] Promoter methylation is suggested to be an early event in carcinogenesis and can be detected in biological fluids in various cancers.[4-7] Body fluids that have been used for cancer screening with methylation markers include urine[8], ejaculates of men[9], salivary rinse,[10], sputum[11], peritoneal fluid[12] and ductal lavage and nipple fluid[13,14] highlighting the potential for application in routine clinical practice.

For CRC, we and others have shown that detection of promoter methylation in fecal DNA holds promise as a CRC prescreening modality.[7,15-20] Genes known to be methylated, detected and studied in tumor derived DNA in blood of CRC patients are *ALX4*[21], *TMEFF2*[22], *CDKN2A (p16)*[23], *CDH4*[24], *NEUROG1*[25], *NGFR*[22], *SEPT9*[22], *MLH1*[26,27], *DAPK*[28], *THBD*[29], SDC2[30], and gene panels consisting of *APC, MLH1* and *HLTF*[27] and *APC*, *MGMT*, *RASSF2A* and *Wif-1*[31]. *The* sensitivity and specificity to detect CRC observed in these studies range from 21% to 86% and from 69% to 100% respectively.

Our objective was to examine promoter methylation of two previously identified stool markers (*NDRG4* and *GATA5*)[15,18] and two novel markers namely *FOXE1* and *SYNE1*[32], as potential biomarkers for the early detection of CRC in blood DNA. *FOXE1* and *SYNE1* were identified as frequently methylated genes using a transcriptome-wide approach to identify genes that are transcriptionally silenced by methylation in CRC.[32] In addition, methylation of *SYNE1* and *FOXE1* has been described in a small cohort of patients with colitis-associated colorectal neoplasia.[33]

Performance of the best combinatorial marker panel was evaluated by quantitative methylation analysis in two large sets of plasma samples from CRC patients and controls. Furthermore, the currently unknown functional role of *SYNE1* and *FOXE1* in CRC was investigated.

Materials and methods

Study population plasma samples

Two hundred and twenty plasma samples were prospectively collected from CRC patients from multiple centers in Germany. Symptomatic patients were screened using colonoscopy and the clinical diagnosis of CRC was confirmed by histology. The trial started in 2007 and recruited patients with all disease stages. Included patients were diagnosed with CRC, had not been treated for CRC prior to blood collection, had not been treated for other malignancies during the previous five years and had surgery planned to assess the UICC stage and the involvement of lymph nodes.

Control blood samples (n=664) were collected from 550 asymptomatic average risk and 134 symptomatic individuals, all without adenomas and/or CRC detected by colonoscopy screening. These individuals were enrolled in a multicentre CRC screening trial in Germany of average risk subjects. Participants underwent primary colonoscopy screening and blood samples were drawn prior to the procedure. Patient characteristics of CRC patients and controls are shown in Supplemental Table S6.1. Plasma samples of CRC patients and controls were randomized and divided in two different sets, one training and one test set as depicted in Supplemental Figure S6.1. Informed consent was obtained from all participants, adhering to ethics guidelines.

Collection and isolation of plasma samples

Nine ml of blood, using 10 ml EDTA Vacutainer™ tubes (BD Vacutainer, BD Hemogard, K2 EDTA spray-dried, Cat#: 367525), were collected using standard venipuncture techniques. Plasma was separated by centrifugation at 1500g for 15 minutes (double spin) within 2 hours of collection. DNA isolation from plasma was performed using the QIAamp Circulating Nucleic Acid Test Kit (Qiagen, Cat#: 55114).

Sodium bisulfite treatment and quantitative methylation-specific PCR

Sodium bisulfite modification was performed using the EpiTect Bisulphite Kit (Qiagen, Cat#59104) according to the manufacturer's instructions. Quantitative methylation-specific PCR on plasma samples was performed on a 7900HT real-time PCR system (Applied Biosystems). For the training set two duplex PCRs *NDRG4_FAM*/*FOXE1_TET* and *GATA5_FAM*/*SYNE1_TET* were performed. Based on the performance of the first duplex reaction, only *SYNE1* and *FOXE1* were retained in the test set (one duplex PCR; *SYNE1_TET/FOXE1_HEX*). The PCR master mix was: 30 µl QuantiTect Multiplex PCR with ROX dye (Qiagen), 10 µl template volume, forward primer (0.28 µM) of both genes, reverse primer (0.28 µM) of both genes and a single-stranded oligonucleotide hybridization probe (0.25 µM) of both genes. The PCR program was: 15 minutes at 95°C; followed by 45 cycles of 30 seconds at 95°C, 30 seconds at 57°C, and 30 seconds

at 72°C; followed by 5 minutes at 72°C. Serially diluted plasmids containing the target sequence were amplified to generate a standard curve against which the unknown samples are quantified by interpolation of their PCR cycle number (Ct value) to the corresponding plasmid copy. Because the yield showed very high inter-subject variability, no direct control for DNA yield was incorporated. This standard curve ruled out any technical or reagent batch related errors in methylation calls. Samples were handled and analyzed in a blinded fashion during storage, DNA isolation, and PCR analysis. Primer and probe sequences are provided in Supplemental Table S6.2.

Cell culture and transfections

Human CRC cell lines were cultured in Dulbecco's MEM (DMEM) (Invitrogen, Breda, the Netherlands) supplemented with 10% heat-inactivated fetal calf serum (FBS; HyClone,Etten-Leur, the Netherlands). Cell lines (RKO, SW480 and HCT116) were purchased from ATCC (LGC standards). No authentication was done by the authors. Full length *FOXE1* cDNA was subcloned into the pIRES-neo3 expression vector (Clonetech Laboratories Inc. CA, USA). GFP-*SYNE1* and GFP-*SYNE1*-KASH (Klarischt) were kindly provided by Dr. Zhang (Department of Medicine, Cambridge, UK).

HCT116 cells were transfected with the Nucleofector Kit V (Amaxa Biosystems, Gaithersburg, MD) and RKO and SW480 cells were transfected using Lipofectamine 2000 Reagent (Invitrogen) according to the manufacturer's protocol. HCT116, RKO and SW480 cells were transfected with a control construct (empty vector) or *FOXE1*-pIRESneo3, GFP-*SYNE1* and GFP-*SYNE1*-KASH, selected for 10 days with G418 (HCT116 and SW480 400 µg/ml; RKO 1 mg/ml).

Colony formation assay

RKO, HCT116 and SW480 CRC cells were transfected in six well-plates with control construct *FOXE1*-pIRESneo3, GFP-*SYNE1* or GFP-*SYNE1*-KASH as described above. The next day, cells were diluted 1:20 and G418 (RKO 1 mg/ml; HCT116, SW480 400 µg/ml) was added. After 14 days of selection, colonies were stained and quantified.

In vitro cell proliferation, migration and invasion assays

HCT116, RKO and SW480 cells were seeded onto 96-well plates (5000 cells/well) and after 96 hours the cultures were pulsed for 6 hours with 0.3 µCi [methyl-^{3}H] thymidine (Amersham Life Science, Roosendaal, the Netherlands) per well. Activity was measured using liquid scintillation. Cell migration and invasion assays were performed using 24-well transwells (8µm pore size) coated with (invasion) or without (migration) matrigel (BD Biosciences, Franklin Lakes, NJ). HCT116, RKO and SW480 cells ($20x10^{4}$) in 1% FBS-DMEM were seeded into the upper chamber, and DMEM containing 20% FBS was

placed in the lower chamber. After 48 hours, cells on the lower surface of the membrane were fixed with methanol and stained with 1% Toluine Blue in 1% borax.

Data analysis

For quantitative methylation-specific PCR analysis, receiver operating characteristics (ROC) curve analysis and the area under the curve (AUC) were determined to define the best markers with highest sensitivity and specificity. *GATA5*, *NDRG4*, *SYNE1* and *FOXE1* promoter methylation was considered positive if copies of any of these genes were detected. The measured methylated copy numbers were used as the basis for the age- and gender-association analyses, hence methylation was implemented as a continuous variable. Age was stratified in two groups, i.e. those patients younger than 65 and those age 65 and above. The Mann-Whitney rank sum test was used to analyze the colony formation, migration and invasion assays.

Results

NDRG4, GATA5, FOXE1 and SYNE1 methylation in blood DNA as a potential biomarker for colorectal cancer detection

The main objective of this study was to investigate the use of *NDRG4*, *GATA5*, *SYNE1* and *FOXE1* promoter methylation as biomarkers for detection of CRC-derived DNA in plasma by determining the optimal classifier in a training set and subsequently confirming this in a test set.

Training set

Using 444 normal control samples and 154 CRC samples of the training set, we evaluated *NDRG4*, *GATA5*, *SYNE1* and *FOXE1* as single marker candidates using ROC curveanalysis. The area under the curve generated for plasma *NDRG4*, *GATA5*, *SYNE1* and *FOXE1* methylation was 0.61 (95%CI=0.57-0.65, Figure 6.1A), 0.59 (95%CI=0.55-0.63, Figure 6.1B), 0.72 (95%CI=0.68-0.75, Figure 6.1C) and 0.70 (95%CI=0.69-0.73, Figure 6.1D), respectively, indicating that *SYNE1* and *FOXE1* have the highest performance as potential biomarkers.

To study the sensitivity and specificity of these markers separately we defined a cutoff of zero. The sensitivity of *NDRG4* methylation in all stages of CRC was 27% (95%CI=20%-34%) with a specificity of 95% (95%CI=93%-97%). The overall sensitivity of *GATA5* methylation was 18% (95%CI=12%-24%) with a specificity of 99% (95%CI=98%-100%). *FOXE1* methylation has a sensitivity of 46% (95%CI=38%-54%) in detecting patients with CRC and a specificity of 93% (95%CI=91%-95%). Finally, *SYNE1* methylation generated a sensitivity of 47% (95% CI=39%-55%) for detection of CRC and a specificity of 96% (95%CI=94%-98%). Early stage CRCs were detected at slightly lower

rates than later stages, a phenomenon that occurred for three of the four markers analyzed (except *SYNE1*), as depicted in (Table 6.1).

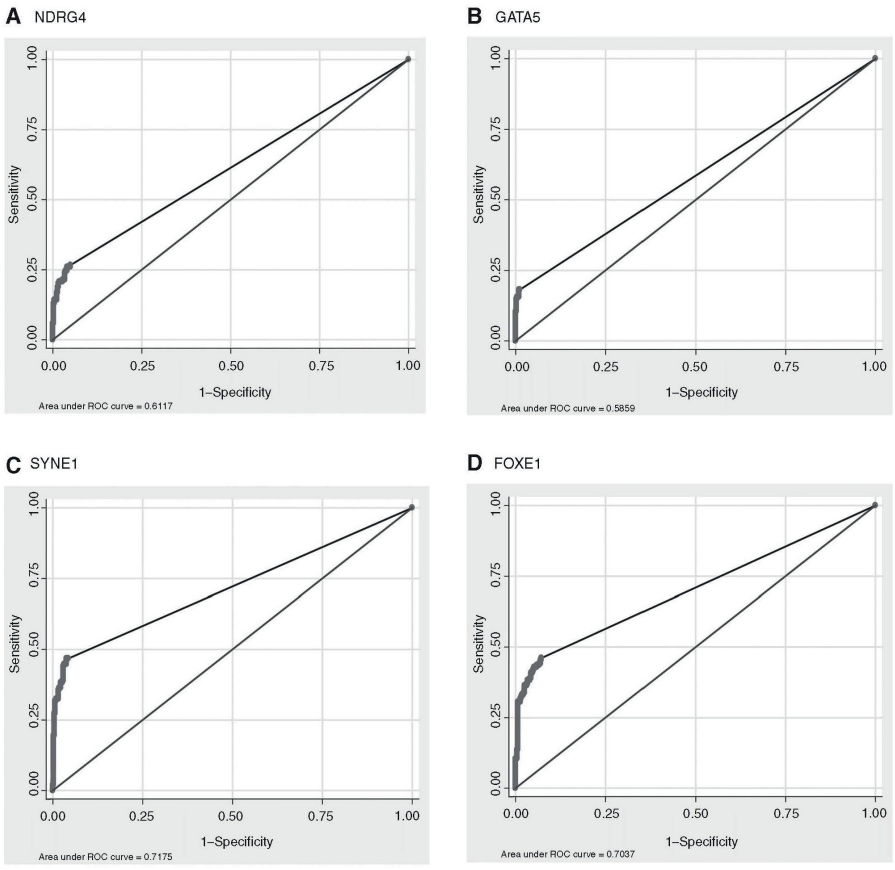

Figure 6.1 **ROC curves.**
Sensitivity and specificity at various cutoff values for the training set to obtain a positive test for the single markers (**A**) *NDRG4*, (**B**) *GATA5*, (**C**) *SYNE1*, and (**D**) *FOXE1* are shown in the ROC curve.

The presence of methylated *NDRG4* (p=0.45, p=0.12) *GATA5* (p=0.60; p=0.13) *FOXE1* (p=0.88; p=0.053) and *SYNE1* (p=0.39; p=0.15) in plasma was not associated with age and gender, respectively. Given the low sensitivity of *GATA5* and *NDRG4* methylation, these markers did not add to the overall sensitivity (data not shown). In addition, combining *SYNE1* and *FOXE1* methylation increased the sensitivity to 56% (95%CI=48%-64%) with a specificity of 90% (95%CI=87%-93%) (Table 6.2).

Table 6.1 Training set plasma performance of single markers.
(**A**) *NDRG4* and *GATA5*, (**B**) *FOXE1* and *SYNE1*.

A

	NDRG4		*GATA5*	
Patient group	Positive	Sensitivity	Positive	Sensitivity
Stage I	7/43	16%	6/43	14%
Stage II	5/44	11%	4/44	9%
Stage III	16/46	35%	8/46	17%
Stage IV	13/21	62%	10/21	48%
Case total	41/154	27%	28/154	18%
		Specificity		Specificity
Control	22/442	95%	5/444	99%

B

	FOXE1		*SYNE1*	
Patient group	Positive	Sensitivity	Positive	Sensitivity
Stage I	15/43	35%	12/43	28%
Stage II	19/44	43%	23/44	52%
Stage III	23/46	50%	21/46	46%
Stage IV	14/21	67%	16/46	76%
Case total	71/154	46%	72/154	47%
		Specificity		Specificity
Control	32/444	93%	19/444	96%

Table 6.2 Training set results combined *FOXE1* and *SYNE1*.

	SYNE1 and *FOXE1*	
Patient group	Positive	Sensitivity
Stage I	18/43	42%
Stage II	25/44	57%
Stage III	27/46	59%
Stage IV	17/21	81%
Case total	87/154	56%
		Specificity
Control	46/444	90%

Test set

To confirm the clinical performance of *SYNE1* and *FOXE1* methylation as a multi-marker panel, we tested plasma samples obtained from 66 CRC patients and 240 healthy controls. Using the same threshold, we observed a sensitivity of 58% (95%CI=46%-70%) and a specificity of 91% (95%CI=80%-100%) for all stages of CRC for *FOXE1* and *SYNE1*, respectively (Table 6.3). Stage I (37% (95%CI=19%-55%)) and stage III (55% (95%CI= 33%-77%)) cancers were detected at lower rates than stage II (87% (95%CI=70%-100%)) and stage IV (100% (95%CI= NA)) cancers (Table 6.3). *SYNE1* and *FOXE1* promoter methylation was not associated with age (*FOXE1*, p=0.36 and *SYNE1*, p=0.38) and gender (*FOXE1* p=0.29 and *SYNE1* p=0.28).

Table 6.3 Test set results combined *FOXE1* and *SYNE1*.

	SYNE1 and *FOXE1*	
Patient group	Positive	Sensitivity
Stage I	10/27	37%
Stage II	13/15	87%
Stage III	11/20	55%
Stage IV	4/4	100%
Case total	38/66	58%
		Specificity
Control	21/240	91%

FOXE1 and SYNE1 overexpression in colorectal cancer cell lines

The functional role of NDRG4 and GATA5 *in vitro* in CRC cell lines has been previously described(15, 18). Currently nothing is known about the role of SYNE1 and FOXE1 in CRC.

To examine whether SYNE1 has tumor suppressor activities in CRC cells, HCT116 and RKO cells were transfected with *SYNE1* expression constructs. Transfection with GFP-*SYNE1* and GFP-*SYNE1*-KASH induced the number of G418-resistant colonies as compared to control transfectants in HCT116 (Figure 6.2A, p=0.001 and p=0.07) and RKO (Figure 6.2B, p=0.13 and p=0.17) cells. Migration (Figure 6.2C (HCT116, p=0.26 and p=0.16) and 2D (RKO, p=0.19 and p=0.07) and invasion (Figure 6.2E (HCT116, p=0.29, and p=0.20) and 2F (RKO, p=0.27 and p=0.03) of *SYNE1* transfected cells was higher than that of control clones. However, although a trend was seen in all cell lines, most results were not statistically significant.

To investigate whether FOXE1 is a putative tumor suppressor in CRC we first performed *in vitro* colony formation assays to determine the effects of full-length *FOXE1* transfected into SW480, HCT116, and RKO cells. Overexpression with *FOXE1* resulted in a significant reduction of G418-resistant colonies in HCT116 and SW480 cells (a nearly 2-fold reduction in HCT116 (Figure 6.3A, p=0.003) and over 8-fold reduction in SW480 cells (Figure 6.3C, p=0.006). No significant effect was observed in RKO cells transfected with *FOXE1* (Figure 6.3B, p=0.577). Overexpression of the *FOXE1* gene had no significant effect on the migration (Figure 6.3D (HCT116, p=0.63), E (RKO, p=0.82), F (SW480, p=0.28)) and invasion (Figure 6.3G HCT116, p=0.79), H (RKO, p=0.92), I (SW480, p=0.38)) of these cells. Taken together, these data suggest that *FOXE1* may have tumor-suppressive effects in human CRC cells.

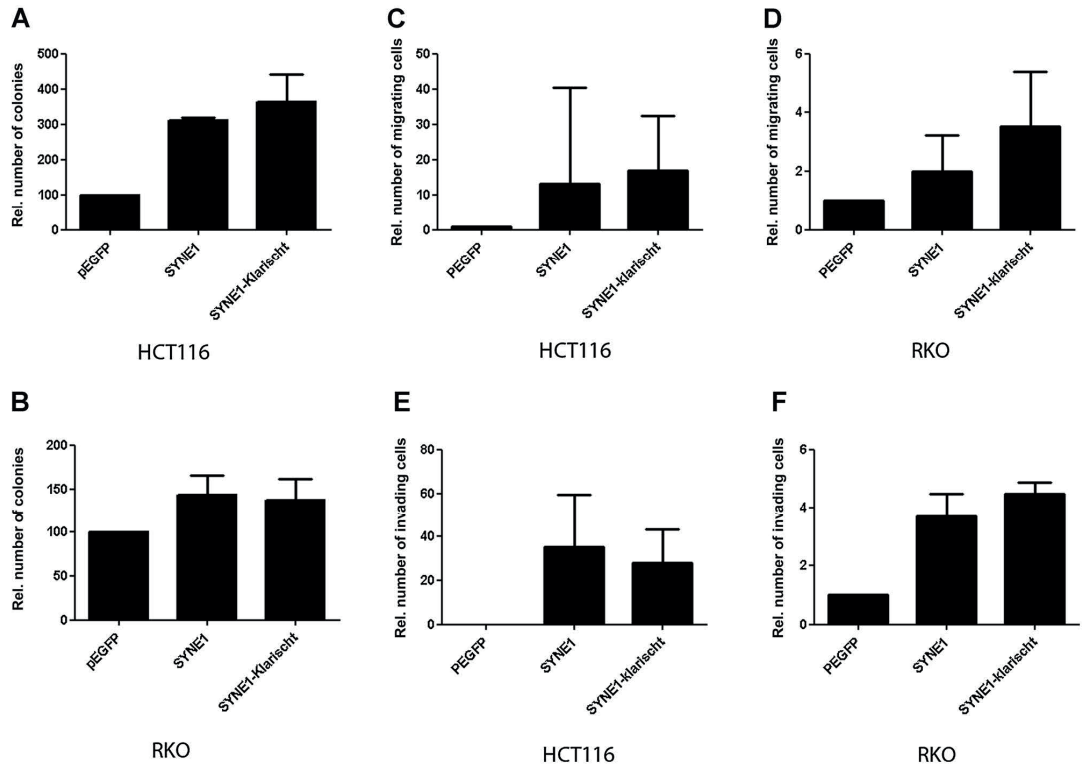

Figure 6.2 **Functional assays *SYNE1*.**
A and **B**, colony formation assay of cells transfectedwith *SYNE1* expression vector or empty vector grown for 2 weeks in medium containing antibiotics. Mean colony numbers relative to control transfectants are plotted (n=3). Error bars, 95% CI. **C** and **D**, invasion assay through Matrigel-coated Transwells. Results represent mean number of *SYNE1*-transfected cells that passed through the Matrigel-coated membranes of the Transwell relative to control cells transfected with empty vector (n=3). **E** and **F**, migration assay. Plotted are the mean numbers of *SYNE1*-transfected cells that migrated through Transwell membranes not coated with Matrigel relative to control cells transfected with empty vector (n=3). Error bars, 95% CI.

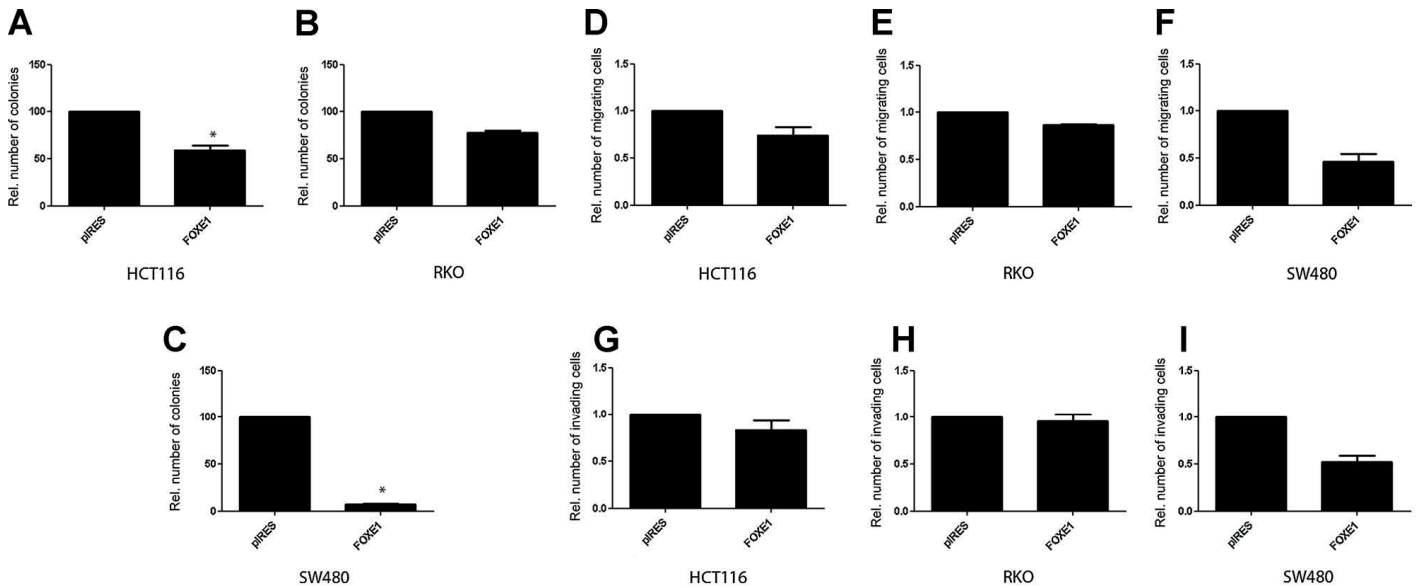

Figure 6.3 Functional assays *FOXE1*.

A to **C**, colony formation assay of cells transfected with *FOXE1* expression vector or empty vector grown for 2 weeks in medium containing antibiotics. Mean colony numbers relative to control transfectants are plotted (n=3). Error bars, 95% CI. **D** to **F**, invasion assay through Matrigel-coated Transwells. Results represent mean number of FOX1-transfected cells that passed through the Matrigel-coated membranes of the Transwell relative to control cells transfected with empty vector (n=3). **G** to **I**, migration assay. Plotted are the mean numbers of *FOX1*-transfected cells that migrated through Transwell membranes not coated with Matrigel relative to control cells transfected with empty vector (n=3). Error bars, 95% CI.

Discussion

We previously published *NDRG4*[18] and *GATA4*/*GATA5*[15] promoter hypermethylation as potential sensitive and specific stool DNA markers and candidate tumor suppressor genes in CRC. The potential of *NDRG4* methylation for the detection of CRC was independently confirmed by Alhquist *et al*. and Imperiale *et al*., who recently develop a multi-marker diagnostic DNA test, including *NDRG4* methylation, to screen for CRC.[19,34] This DNA test has shown to perform better compared to analyzing FIT alone. However, implementation of this test might be hampered in population-based screening programs due to the requirement of a complete stool sample affecting compliance and cost-effectiveness. Despite the high potential of these markers in stool DNA, we hypothesized that a blood-based test for CRC detection, not depending on stool sampling, has the potential for better patient compliance and is better suited for systems without programmatic screening. Therefore we investigated the performance of our previously reported stool methylation markers (*NDRG4* and *GATA5*) and two other promising methylation markers (*SYNE1* and *FOXE1*) in blood. *FOXE1* and *SYNE1* were identified as frequently methylated genes in a transcriptome-wide approach to detect transcriptionally silenced genes by promoter CpG island methylation in CRC.[32]

FOXE1 is a transcription factor which is characterized by a distinct forkhead domain, and plays a crucial role in thyroid morphogenesis. In thyroid cancer *FOXE1* is upregulated[35], although most cancers have a decreased expression of *FOXE1* often due to promoter CpG island methylation as has been described in breast cancer[36], pancreatic cancer[37], and in cutaneous squamous cell carcinoma[38]. Furthermore, *FOXE1* has been reported as a sensitive (82%) and specific (98.5%) methylation marker for pancreatic cancer in pancreatic juice of patients.[37]

We and others described DNA methylation of *SYNE1* in CRC, suggesting a tumor suppressor function in CRC.[32,39] *SYNE1* promoter methylation and a possible tumor suppressor function have been reported in lung cancer.[40] However the *in vitro* studies in this study show that *SYNE1* overexpression in CRC cell lines induces cell proliferation, migration and invasion indicating an oncogenic role for *SYNE1* in CRC. Interestingly, in line with our functional data, gene expression data obtained from GEO from primary normal colon and colon cancer showed no decrease of *SYNE1* expression in CRC tissues (data not shown), speculating that promoter methylation of *SYNE1* in CRC does not result in downregulation of this gene.

We have studied the role of NDRG4 in CRC and found that NDRG4 significantly inhibited colony formation, cell proliferation, and invasion and is frequently methylated (70% and 86% in two independent series) in CRC tissue samples.[18] In parallel, we have published the potential tumor suppressor role of GATA5 in CRC, which inhibited colony formation, cell growth, migration, invasion, and anchorage-independent growth *in vitro*. Furthermore, promoter methylation of *GATA5* was observed (79% and 74% in two independent series) in CRC tissues.[15] *GATA5* promoter methylation has been reported

in other tumor types as well, such as small cell lung cancer, pancreatic, esophageal, ovarian and gastric cancer.[41-45] In the present paper, we investigated the performance of *NDRG4*, *GATA5*, *SYNE1* and *FOXE1* as methylation markers in blood DNA. Interestingly, although *NDRG4* performed well as a methylation marker in stool (sensitivity of 61% and 53% for the training and test set, respectively), the detection rate in blood for CRC was only 27%. Molecular pathways and cellular mechanisms that underlie multistage processes of metastasis, including tumor invasion, tumor-cell dissemination through the bloodstream or the lymphatic system, colonization of distant organs and, outgrowth of metastases have been characterized.[46] A difference in gene expression at the invasive front compared to the central area of the tumor and the luminal part of the tumor has been described for beta-catenin.[47] This heterogeneity in gene expression probably indicates heterogeneity in the underlying mechanisms that are regulating gene expression. We therefore hypothesize that molecular or morphological characteristics at the invasive front area of CRC, a part of the tumor that probably also easily invades blood vessels, may be different from the luminal part of the CRC that is shedding tumor cells in stool. An alternative explanation for differences in sensitivity in blood or fecal DNA could be the DNA isolation procedure. Diehl *et al*. demonstrated that the majority of mutated *APC* sequences in patients with CRC were detected in smaller size fragments of circulating DNA, whereas larger fragments tended to be wild-type.[48] Therefore, improvement of isolation protocols of blood DNA, could, in theory, yield a blood DNA test for *NDRG4* with higher sensitivity.

We here describe a biomarker panel of two genes able to discriminate CRC from non-CRC in plasma with a sensitivity of 56% and 58% and a specificity of 90% and 91%. Of all blood-based DNA markers, plasma septin-9 has been studied most extensively. Test sensitivity varies widely for both CRC (from 48% to 90%) and adenomas (from 11% to 29%) while specificity was more consistent ranging from 86 to 97%.[49] Septin-9 was also evaluated in a screening setting, where sensitivity and specificity were 50% and 91%, respectively, for CRC patients when two aliquots were tested and 64% and 88% respectively with three samples tested.[49] Further research is necessary to investigate whether the sensitivity of *SYNE1* and *FOXE1* can be improved using different aliquots and whether it will be effective and cost-effective compared with no screening and compared to other screening test available. In addition, these data need to be validated in a large prospective CRC screening study and should be combined with other non-invasive screening tests such as genetic or epigenetic DNA markers or the fecal immunochemical test (FIT) to enhance sensitivity. Furthermore, sample collection, DNA isolation, bisulfite conversion and assay sensitivity should be optimized. Detecting minute amounts of tumor DNA in blood, especially in the early stage tumors, and defining the specificity of a DNA methylation marker in blood that contain DNA from many sources in the body are challenges for future research.

References

1. Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. N Engl J Med 1993;328:1365-1371.
2. Wee CC, McCarthy EP, Phillips RS. Factors associated with colon cancer screening: the role of patient factors and physician counseling. Prev Med 2005;41:23-29.
3. Jones PA, Baylin SB. The epigenomics of cancer. Cell 2007;128:683-692.
4. Laird PW. The power and the promise of DNA methylation markers. Nat Rev Cancer 2003;3:253-266.
5. Hoque MO, Begum S, Topaloglu O, Chatterjee A, Rosenbaum E, Van Criekinge W, et al. Quantitation of promoter methylation of multiple genes in urine DNA and bladder cancer detection. J Natl Cancer Inst 2006;98:996-1004.
6. Jin Z, Olaru A, Yang J, Sato F, Cheng Y, Kan T, et al. Hypermethylation of tachykinin-1 is a potential biomarker in human esophageal cancer. Clin Cancer Res 2007;13:6293-6300.
7. Muller HM, Oberwalder M, Fiegl H, Morandell M, Goebel G, Zitt M, et al. Methylation changes in faecal DNA: a marker for colorectal cancer screening? Lancet 2004;363:1283-1285.
8. Yu J, Zhu T, Wang Z, Zhang H, Qian Z, Xu H, et al. A novel set of DNA methylation markers in urine sediments for sensitive/specific detection of bladder cancer. Clin Cancer Res 2007;13:7296-7304.
9. Goessl C, Krause H, Muller M, Heicappell R, Schrader M, Sachsinger J, et al. Fluorescent methylation-specific polymerase chain reaction for DNA-based detection of prostate cancer in bodily fluids. Cancer Res 2000;60:5941-5945.
10. Rosas SL, Koch W, da Costa Carvalho MG, Wu L, Califano J, Westra W, et al. Promoter hypermethylation patterns of p16, O6-methylguanine-DNA-methyltransferase, and death-associated protein kinase in tumors and saliva of head and neck cancer patients. Cancer Res 2001;61:939-942.
11. Machida EO, Brock MV, Hooker CM, Nakayama J, Ishida A, Amano J, et al. Hypermethylation of ASC/TMS1 is a sputum marker for late-stage lung cancer. Cancer Res 2006;66:6210-6218.
12. Muller HM, Millinger S, Fiegl H, Goebel G, Ivarsson L, Widschwendter A, et al. Analysis of methylated genes in peritoneal fluids of ovarian cancer patients: a new prognostic tool. Clin Chem 2004;50: 2171-2173.
13. Evron E, Dooley WC, Umbricht CB, Rosenthal D, Sacchi N, Gabrielson E, et al. Detection of breast cancer cells in ductal lavage fluid by methylation-specific PCR. Lancet 2001;357:1335-1336.
14. Krassenstein R, Sauter E, Dulaimi E, Battagli C, Ehya H, Klein-Szanto A, et al. Detection of breast cancer in nipple aspirate fluid by CpG island hypermethylation. Clin Cancer Res 2004;10:28-32.
15. Hellebrekers DM, Lentjes MH, van den Bosch SM, Melotte V, Wouters KA, Daenen KL, et al. GATA4 and GATA5 are potential tumor suppressors and biomarkers in colorectal cancer. Clin Cancer Res 2009;15:3990-3997.
16. Chen WD, Han ZJ, Skoletsky J, Olson J, Sah J, Myeroff L, et al. Detection in fecal DNA of colon cancer-specific methylation of the nonexpressed vimentin gene. J Natl Cancer Inst 2005;97:1124-1132.
17. Lenhard K, Bommer GT, Asutay S, Schauer R, Brabletz T, Goke B, et al. Analysis of promoter methylation in stool: a novel method for the detection of colorectal cancer. Clin Gastroenterol Hepatol 2005;3: 142-149.
18. Melotte V, Lentjes MH, van den Bosch SM, Hellebrekers DM, de Hoon JP, Wouters KA, et al. N-Myc downstream-regulated gene 4 (NDRG4): a candidate tumor suppressor gene and potential biomarker for colorectal cancer. J Natl Cancer Inst 2009;101:916-927.
19. Ahlquist DA, Zou H, Domanico M, Mahoney DW, Yab TC, Taylor WR, et al. Next-generation stool DNA test accurately detects colorectal cancer and large adenomas. Gastroenterology 2012;142:248-256.
20. Glockner SC, Dhir M, Yi JM, McGarvey KE, Van Neste L, Louwagie J, et al. Methylation of TFPI2 in stool DNA: a potential novel biomarker for the detection of colorectal cancer. Cancer Res 2009;69: 4691-4699.
21. Ebert MP, Model F, Mooney S, Hale K, Lograsso J, Tonnes-Priddy L, et al. Aristaless-like homeobox-4 gene methylation is a potential marker for colorectal adenocarcinomas. Gastroenterology 2006;131: 1418-1430.

22. Lofton-Day C, Model F, Devos T, Tetzner R, Distler J, Schuster M, et al. DNA methylation biomarkers for blood-based colorectal cancer screening. Clin Chem 2008;54:414-423.
23. Zou H, Yu B, Zhao R, Wang Z, Cang H, Li D, et al. Detection of aberrant p16 methylation in the serum of colorectal cancer patients. Zhonghua Yu Fang Yi Xue Za Zhi 2002;36:499-501.
24. Miotto E, Sabbioni S, Veronese A, Calin GA, Gullini S, Liboni A, et al. Frequent aberrant methylation of the CDH4 gene promoter in human colorectal and gastric cancer. Cancer Res 2004;64:8156-8159.
25. Herbst A, Rahmig K, Stieber P, Philipp A, Jung A, Ofner A, et al. Methylation of NEUROG1 in serum is a sensitive marker for the detection of early colorectal cancer. Am J Gastroenterol 2011;106:1110-1118.
26. Grady WM, Rajput A, Lutterbaugh JD, Markowitz SD. Detection of aberrantly methylated hMLH1 promoter DNA in the serum of patients with microsatellite unstable colon cancer. Cancer Res 2001;61: 900-902.
27. Leung WK, To KF, Man EP, Chan MW, Bai AH, Hui AJ, et al. Quantitative detection of promoter hypermethylation in multiple genes in the serum of patients with colorectal cancer. Am J Gastroenterol 2005;100:2274-2279.
28. Yamaguchi S, Asao T, Nakamura J, Ide M, Kuwano H. High frequency of DAP-kinase gene promoter methylation in colorectal cancer specimens and its identification in serum. Cancer Lett 2003;194: 99-105.
29. Lange CP, Campan M, Hinoue T, Schmitz RF, van der Meulen-de Jong AE, Slingerland H, et al. Genome-scale discovery of DNA-methylation biomarkers for blood-based detection of colorectal cancer. PLoS One 2012;7:e50266.
30. Oh T, Kim N, Moon Y, Kim MS, Hoehn BD, Park CH, et al. Genome-wide identification and validation of a novel methylation biomarker, SDC2, for blood-based detection of colorectal cancer. The Journal of molecular diagnostics : J Mol Diagn 2013;15:498-507.
31. Lee BB, Lee EJ, Jung EH, Chun HK, Chang DK, Song SY, et al. Aberrant methylation of APC, MGMT, RASSF2A, and Wif-1 genes in plasma as a biomarker for early detection of colorectal cancer. Clin Cancer Res 2009;15:6185-191.
32. Schuebel KE, Chen W, Cope L, Glockner SC, Suzuki H, Yi JM, et al. Comparing the DNA hypermethylome with gene mutations in human colorectal cancer. PLoS Genet 2007;3:1709-1723.
33. Papadia C, Louwagie J, Del Rio P, Grooteclaes M, Coruzzi A, Montana C, et al. FOXE1 and SYNE1 genes hypermethylation panel as promising biomarker in colitis-associated colorectal neoplasia. Inflamm Bowel Dis 2014;20:271-277.
34. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, et al. Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med 2014;370:1287-1297.
35. Nonaka D, Tang Y, Chiriboga L, Rivera M, Ghossein R. Diagnostic utility of thyroid transcription factors Pax8 and TTF-2 (FoxE1) in thyroid epithelial neoplasms. Mod Pathol 2008;21:192-200.
36. Weisenberger DJ, Trinh BN, Campan M, Sharma S, Long TI, Ananthnarayan S, et al. DNA methylation analysis by digital bisulfite genomic sequencing and digital MethyLight. Nucleic Acids Res 2008;36: 4689-4698.
37. Matsubayashi H, Canto M, Sato N, Klein A, Abe T, Yamashita K, et al. DNA methylation alterations in the pancreatic juice of patients with suspected pancreatic disease. Cancer research. 2006;66:1208-17.
38. Venza I, Visalli M, Tripodo B, De Grazia G, Loddo S, Teti D, et al. FOXE1 is a target for aberrant methylation in cutaneous squamous cell carcinoma. Br J Dermatol 2010;162(5):1093-1097.
39. Mokarram P, Kumar K, Brim H, Naghibalhossaini F, Saberi-firoozi M, Nouraie M, et al. Distinct high-profile methylated genes in colorectal cancer. PLoS One 2009;4:e7012.
40. Tessema M, Willink R, Do K, Yu YY, Yu W, Machida EO, et al. Promoter methylation of genes in and around the candidate lung cancer susceptibility locus 6q23-25. Cancer Res 2008;68:1707-1714.
41. De Jong WK, Verpooten GF, Kramer H, Louwagie J, Groen HJ. Promoter methylation primarily occurs in tumor cells of patients with non-small cell lung cancer. Anticancer Res 2009;29:363-369.
42. Fu B, Guo M, Wang S, Campagna D, Luo M, Herman JG, et al. Evaluation of GATA-4 and GATA-5 methylation profiles in human pancreatic cancers indicate promoter methylation patterns distinct from other human tumor types. Cancer Biol Ther 2007;6:1546-1552.
43. Guo M, House MG, Akiyama Y, Qi Y, Capagna D, Harmon J, et al. Hypermethylation of the GATA gene family in esophageal cancer. Int J Cancer 2006;119:2078-2083.

44. Wakana K, Akiyama Y, Aso T, Yuasa Y. Involvement of GATA-4/-5 transcription factors in ovarian carcinogenesis. Cancer Lett 2006;241:281-288.
45. Akiyama Y, Watkins N, Suzuki H, Jair KW, van Engeland M, Esteller M, et al. GATA-4 and GATA-5 transcription factor genes and potential downstream antitumor target genes are epigenetically silenced in colorectal and gastric cancer. Mol Cell Biol 2003;23:8429-8439.
46. Christofori G. New signals from the invasive front. Nature 2006;441:444-450.
47. Brabletz T, Jung A, Hermann K, Gunther K, Hohenberger W, Kirchner T. Nuclear overexpression of the oncoprotein beta-catenin in colorectal cancer is localized predominantly at the invasion front. Pathol Res Pract 1998;194:701-704.
48. Diehl F, Li M, Dressman D, He Y, Shen D, Szabo S, et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc Natl Acad Sci U S A 2005;102:16368-16373.
49. Imperiale TF. Noninvasive screening tests for colorectal cancer. Dig Dis 2012;30 Suppl 2:16-26.

Supplementary figure and tables

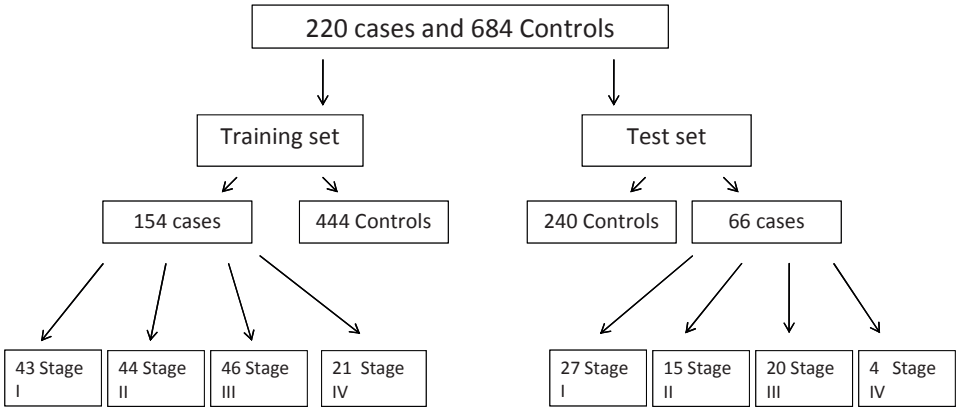

Figure S6.1 Study population.
220 plasma samples from CRC patients were obtained from a multicenter prospective trial in Germany. Control blood samples (n=664) were collected from 550 asymptomatic average risk and 134 symptomatic individuals, all without adenomas and/or CRC detected by colonoscopy screening. The subjects were randomized and divided into the training and test cohorts so that the training cohort consisted of roughly 70% of all subjects while maintaining a prevalence for cancer of approximately 25%.

Table S6.1 Patient characteristics of training and test studies.

Diagnosis Group	Description	Gender		Age	
		Male	Female	≤65	>65
Training set	Stage I	23/43	20/43	8/43	35/43
	Stage II	29/44	15/44	14/44	30/44
	Stage III	21/46	25/46	11/46	35/46
	Stage IV	11/21	9/21 (1 NA)	11/21	10/21
	Control	202/444	242/444	253/444	191/444
Test set	Stage I	20/27	7/27	9/27	18/27
	Stage II	9/15	6/15	2/15	13/15
	Stage III	8/20	12/20	10/20	10/20
	Stage IV	1/4	3/4	3/4	1/4
	Control			105/240	93/240 (42 NA)

Table S6.2 Primer and beacon sequences for multiplex assays.

GATA5	forward primer	5'- AGT TCG TTT TTA GGT TAG TTT TCG GC - 3'
	reverse primer	5'- CCA ATA CAA CTA AAC GAA CGA ACC G - 3'
	Beacon	5'- FAM- CGACATGCGTAGGGAGGTAGAGGGTTCGGGATTCGTAGCATGTCG –BHQ1-3'
NDRG4	forward primer	5'- GTATTTTAGTCGCGTAGAAGGC - 3'
	reverse primer	5'- AATTTAACGAATATAAACGCTCGAC - 3'
	Beacon	5'- FAM-CGACATGCCCGAACGAACCGCGATCCCTGCATGTCG-BHQ1-3’
SYNE1	forward primer	5’- GTTGGGTTTTCGTAGTTTTGTAGATCGC – 3’
	reverse primer	5’– CTACGCCCAAACTCGACG – 3’
	Beacon	5'- TET - CGACATGCCCCGCCCTATCGCCGAAATCGCATGTCG –BHQ1-3'
FOXE1	forward primer	5’- TTTGTTCGTTTTTCGATTGTTC -3’
	reverse Primer	5’- TAACGCTATAAAACTCCTACCGC -3’
	Beacon	5’- TET and HEX- CGTCTCGTCGGGGTTCGGGCGTATTTTTTTTAGGTAGGCGAGACG –BHQ1-3’

Chapter 7

General discussion

With over 1,2 million newly diagnosed patients and more than 650,000 deaths annually, colorectal cancer (CRC) is both the third most common cancer, as well as the third highest cause of cancer related death in the developed world.[1] An aging population will cause these numbers to further increase, even if age-specific rates remain constant and mortality rates start to improve. CRC is, in most cases, a curable disease when diagnosed at the precursor stage or at an early stage. Indeed a decrease in CRC incidence after detection and removal of precursor lesions using flexible sigmoidoscopy has been shown.[2] Five year survival rates in patients with localized (stage I) disease are over 90% and less than 20% in patients with metastatic (stage IV) disease.[3-5] However, the majority of cases are diagnosed at a later stage when the cancer has already spread and metastasized. This underscores the need for early detection of CRC. Although some countries, e.g. Greece and Poland, use the gold standard for detection of colorectal neoplasia, colonoscopy, as a screening tool, most CRC screening programs (including the Dutch CRC screening program) have introduced a first-line fecal occult blood test (guaiac-based fecal occult blood test; gFOBT, or fecal immunochemical test; FIT) which, upon positive, will be followed by colonoscopy.[6] Although both methods are noninvasive, cheap and easy to use, the performance of the FIT test still can be improved.[7-14] Non-invasive analysis of cancer-specific molecular alterations has been considered an addition or alternative to FOBT-based tests.[15]

In this thesis we identified and evaluated novel molecular markers for non-invasive detection of CRC.

Stool DNA markers for the early detection of CRC

In chapter 2, we examined promoter methylation of *GATA4* in stool DNA in two independent series of CRC patients using a quantitative methylation-specific PCR (qMSP) assay. The first patient series showed a sensitivity of 71% to detect CRC and specificity of 84% and validation to confirm the test performance in a second patient series yielded a sensitivity of 51% and specificity of 93%. Although the results of these two small pilot studies are promising, further development of *GATA4* as a biomarker for early detection of CRC is dispensable as we identified a biomarker with a better performance, namely *NDRG4*, described in chapter 4. However*, GATA4* and *GATA5* are not only interesting biomarkers because of the relatively high CRC detection rate but also since their potential role in the CRC biology. A tumor suppressor function of GATA4 and GATA5 proteins was suggested and confirmed using different *in vitro* studies in chapter 2. However, the downstream target genes by which GATA4 and GATA5 exhibit these tumor suppressor features are not yet described. Affymetrix oligonucleotide microarray analysis (Human Genome U133 Plus 2.0 GeneChip Array) was performed to search for genes differentially expressed between CRC cells transfected with *GATA4* or *GATA5* expression constructs as compared with control transfectants. The most

upregulated gene in *GATA4* as well as in *GATA5* transfectants was the *plasminogen activator inhibitor* (*PAI-2* or *SerpinB2*). This gene has a (A/T)GATA(A/G) sequence in intron 1 which is recognized by GATA proteins and showed significant upregulation in validation experiments using *GATA4* and *GATA5* transfected CRC cell lines (data not published). PAI-2 is a component of the plasminogen/plasmin system (PPS), one of the main protease systems involved in tumor cell invasion and metastasis.[16] Interestingly, overexpression of PAI-2 in human melanoma cells results in inhibited spontaneous metastasis in scid/scid mice, suggesting a tumor suppressor role in cancer.[17] Additionally, PAI-2 is described as a prognostic marker in which patients with high PAI-2 expression levels tend to have a favorable prognosis in lung[18,19], breast[20] and ovarian[21] cancer. However, in other cancer types such as endometrial[22] and colorectal cancer[23], high PAI-2 expression was correlated with a poor prognosis. Recently, Valiente et al. showed that plasminogen from reactive brain stroma, which inhibits metastatic invasion, is suppressed by PAI-2 and thus leads to initiation of brain metastasis.[24] More studies towards unraveling the function of GATA4 and GATA5 are required to fully understand their role in CRC.

In chapter 4, we described the discovery of the *NDRG4* gene as a biomarker for CRC detection. *NDRG4* promoter methylation occurs frequently in CRC and can be detected in stool with detection rates of 61% in CRC patients in a training set and 53% in the validation set with specificities of 91% and 100%, respectively. However, this study should be considered as a pilot study based on small numbers of patients in a selected set of stool samples. Nevertheless, the clinical potential of *NDRG4* methylation to detect CRC was soon confirmed by other independent studies.[25,26] First in a blinded, multicenter, case-control study stool multiple DNA markers were analyzed on archived stool samples. The multimarker panel included quantitative detection of *KRAS* mutations, and aberrant promoter methylation of *vimentin*, *NDRG4*, *BMP3* and *TFPI2*. Combining the training and test set, the sensitivity of CRC detection reached a level of 85% at a cutoff specificity of 90%. Analysis of the contribution of the individual panel markers showed that *NDRG4* performs best, with an area under the ROC curve of 0.75.[25] At the same specificity cutoff of 90%, adenomas with a size of more than 1 cm were in this study detected in 63% of the cases.[25] Based on these data, the marker panel was limited to *KRAS* mutations, and promoter methylation detection of *NDRG4* and *BMP3*, plus a hemoglobin immunoassay which was called Cologuard®. Next, these results were validated in a large prospective study of 9989 asymptomatic individuals with an average risk for CRC. Cologuard® reached sensitivities of 92% and 42% for detecting CRC and advanced precursor lesions respectively, which is significantly higher when compared to the FIT (which obtained sensitivities of 74% and 24% for detecting CRC and advanced lesions, respectively). Surprisingly, Cologuard® yielded more false positive results compared to the FIT, with specificities of 90% and 96%, respectively.[26] Two thirds of the false positive findings were associated with non-advanced polyps,

raising the question whether these results should be considered as true false positive or high risk precursor lesions that warrant surveillance colonoscopy.[27] Moreover, it is likely that in the future recognition of advanced lesions which will progress to invasive cancer will depend more on molecular aberrations than classical morphological features, such as size and histology.[28,29] The next step in the translation of *NDRG4* to the CRC screening practice was the approval of Cologuard® (Exact Sciences) by the United States Food and Drug Administration (FDA) to screen average risk individuals ≥50 years old for CRC.[30] Additionally, the costs of the Cologuard® are covered by many health insurances in the United States of America.

Although *NDRG4* has been incorporated in a commercial molecular stool test for early detection of CRC, knowledge encompassing the biological role of NDRG4 in the gastrointestinal tract is scarce. Overexpression of NDRG4 in human CRC cell lines revealed several tumor suppressive features as described in chapter 4. Compared to control cells transfected with an empty vector, the NDRG4 transfectants showed reduced colony formation, proliferation and invasion. Next, we explored the expression profile using immunohistochemical and immunofluorescent analysis of human and mouse tissue (chapter 5). In addition to the already described expression in the central nervous system, we observed NDRG4 protein expression in the peripheral nervous system throughout the murine body, including the submucosal (Meissner's) plexus and myenteric (Auerbach's) plexus of the enteric nervous system. Moreover, using whole-mount murine intestinal preparations, we found that NDRG4 protein expression was restricted to enteric neurons, as NDRG4 positive cells were always labeled for the pan-neuronal marker HuCD, but never colocalized with the glial cell marker GFAP. Surprisingly, because of its potential as a biomarker for CRC, no NDRG4 expression was found in the epithelial cells of the gut. It is known that a tumor is dependent on its microenvironment, and that sub-epithelial cells are major regulators in the development/progression of epithelial tumors. While some cells, like endothelial or immune cells, are extensively studied and show promising results as therapeutic targets, the putative role of the enteric neurons in CRC development or progression has not yet been studied. This is surprising since more and more research is being performed on the role of nerve cells in tumorigenesis. It has been described that tumor cells secrete neurotrophic factors thereby initiating their own innervation.[31] In addition, there are several lines of evidence showing that a higher expression of nerve cell markers in the tumor is correlated with a poorer outcome of cancer disease.[32] Because of these data it has been postulated that tumors are able to stimulate their own innervation in a process similar to angiogenesis and lymphangiogenesis.[33] The enteric neurons can also regulate intestinal epithelial cells via the secretion of different neuromediators and it has been shown that alterations in the enteric neurons can lead to different intestinal diseases. However, how these neuromediators may serve as positive or negative epithelial growth signals for CRC development has not yet been investigated. Therefore, more experimental evidence is required to clarify whether

NDRG4, which is specifically expressed in the enteric neurons, might have a role in the development and/or progression from normal gut epithelial cells to malignant cells giving rise to CRC.

Blood or feces?

Although we have identified biomarkers in tumor-derived DNA in the stool for early detection of CRC, stool testing is not optimal when considering patient compliance and laboratory logistics. Many blood-based tests for CRC detection are developed which seem to be promising screening tools due to the high patients acceptance and accessibility next to less transport and storage burden. Therefore we analyzed the performance of *NDRG4* and *GATA5*, and two other novel potential biomarkers *SYNE1* and *FOXE1* promoter hypermethylation in blood (chapter 6). In contrast to the performance of *GATA5* and *NDRG4* methylation in stool-based assays, the sensitivity to detect CRC in plasma was substantially lower, 18% and 27% respectively. *SYNE1* and *FOXE1* showed higher methylation frequencies in plasma of 47% and 46% with specificities of 96% and 93%, respectively. When combining *SYNE1* and *FOXE1* in a panel, the sensitivity increases to 58% with a slightly decreased specificity of 91%. Calculating the sensitivity per stage, shows a clear stage dependency with sensitivities ranging from 37% in stage I to 100% in stage IV CRC. This highlights the difficulties to detect localized disease by blood-based assays as blood-borne spread is needed in order to retrieve tumor-derived DNA in plasma or serum which usually occurs at a higher stage of the tumor development. One of the most extensively studied methylated genes in serum of CRC patients is the *Septin 9* (*SEPT9*) gene, which achieved a sensitivity of 48% with a range of 35-77% for stage I-IV, and specificity of 92%. Only 11% of advanced adenomas could be detected, indicating that the detection of precursor lesions with blood-based assays is not reliable yet.[34] Several studies described an improved detection rate of 77-87% and specificities of 68-92% when combining a variety of DNA-methylation markers in blood.[35-38] Hypothetically, combining methylation detection of *SEPT9* with *SYNE1*, *FOXE1* and *NDRG4* could lead to a higher sensitivity. However, when the plasma test for methylated *SEPT9* is compared to the stool-based Cologuard® test plus methylated *vimentin* and *TFPI2* (PreGen-Plus test), a higher accuracy for the stool DNA test than the plasma *SEPT9* test is obtained with sensitivities of 87% and 60% respectively.[39] Together at this moment stool-based DNA tests seemed to have more potential than blood-based assays in the context of screening for CRC and its precursor lesions.

Optimization of diagnostic CRC biomarker tests

NDRG4 as a biomarker (potentially within the Cologuard®) is promising and should be further developed through improvement of the clinical and analytical sensitivity. The clinical sensitivity should be increased by identification of novel/complementary markers by for example methylation profiling of stool DNA using Methyl Binding Domain (MBD)-affinity capture followed by massive parallel sequencing of FIT-negative tumors. The Cancer Genome Atlas (TCGA) project has offered a valuable catalogue of genomic and epigenomic aberrations which can be used for the identification of yet unknown biomarkers or finding biomarkers that can complement *NDRG4* and the other Cologuard® markers.

Additionally, future studies should focus on enhancement of the analytical sensitivity by developing more accurate and sensitive techniques to isolate DNA from stool or blood and detect molecular alterations. Several factors influence the obtained amount of DNA and its quality, such as sample collection and storage buffers. In order to increase the yield of good quality DNA, these factors should be optimized. A promising development is the combined DNA isolation, bisulfite conversion and methylation detection method called methylation on beads (MOB) in which these processes are incorporated within a single tube through use of silica superparamagnetic beads (SSBs). Bailey et al. showed a 6.6-fold increased median pre-PCR yield compared to conventional techniques when evaluating *CDKN2A* promoter methylation in serum DNA of lung cancer patients.[40] Even more promising is combining MOB with a sensitive methylation marker assay such as the methylation-specific quantum dot fluorescence resonance energy transfer (MS-qFRET). A high sensitivity in detection of methylation using the MS-qFRET method has already been shown in patient sputum samples that contain low concentrations of methylated DNA of the *PYCARD, CDKN2B*, and *CDKN2A* genes, which normally would require a nested PCR approach.[41,42]

Since the first report of cancer biomarkers, numerous scientific papers have been published regarding the identification of putative biomarkers for early detection of cancer, and prediction of prognosis or response to therapy. Despite this large field of research, less than 1% of the published cancer biomarkers has been incorporated into clinical practice. Reasons for the lack of implementation are: insufficient power in marker identification studies leading to false positive results and lack of independent replication in sufficiently powered validation studies.[43,44] Although analysis of several biomarkers shows promising results, lack of clinical applicable assays and/or cost effectiveness also impede incorporation into population-based screening. Evaluation of a promising screening test requires a step-wise approach before acceptance into a screening program can be obtained. First retrospective studies are necessary to ascertain the ability to discriminate between CRC and normal controls. Next, the aimed detection of early neoplasms should be established using prospective clinical studies

testing an asymptomatic population. The third phase encompasses determination of the test characteristics and acceptability in a single round of screening. In the last phase multiple rounds of screening have to be evaluated to estimate the impact of the screening test on reducing cancer mortality and the cost-effectiveness compared to existing tests, such as gFOBT, FIT or flexible sigmoidoscopy.[45]

In conclusion, *NDRG4* is a promising marker for early detection of CRC in stool and should be developed further to incorporate this marker in existing first-line pre-tests for CRC screening such as FIT. Investigating the function of NDRG4 will improve more insights in its role in CRC carcinogenesis.

References

1. Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. *CA Cancer J Clin* 2015;65(2):87-108.
2. Atkin WS, Edwards R, Kralj-Hans I, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *Lancet* 2010:375(9726):1624-1633.
3. Siegel, R., Naishadham, D., and Jemal, A., Cancer statistics, 2012. *CA Cancer J Clin* 2012;62(1):10-29.
4. Kopetz S, Chang GJ, Overman MJ, et al. Improved survival in metastatic colorectal cancer is associated with adoption of hepatic resection and improved chemotherapy. *J Clin Oncol* 2009;27(22):3677-3683.
5. Gunderson LL, Jessup JM, Sargent DJ, et al. Revised TN categorization for colon cancer based on national survival outcomes data. *J Clin Oncol* 2010;28(2):264-271.
6. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. *Gut* 2015;64(10):1637-1649.
7. Graser A, Stieber P, Nagel D, et al. Comparison of CT colonography, colonoscopy, sigmoidoscopy and faecal occult blood tests for the detection of advanced adenoma in an average risk population. *Gut* 2009;58(2):241-248.
8. Brenner H, Tao S, Superior diagnostic performance of faecal immunochemical tests for haemoglobin in a head-to-head comparison with guaiac based faecal occult blood test among 2235 participants of screening colonoscopy. *Eur J Cancer* 2013;49(14):3049-3054.
9. Park DI, Ryu S, Kim YH, et al. Comparison of guaiac-based and quantitative immunochemical fecal occult blood testing in a population at average risk undergoing colorectal cancer screening. *Am J Gastroenterol* 2010;105(9):2017-2025.
10. Chiu HM, Lee YC, Tu CH, et al. Association between early stage colon neoplasms and false-negative results from the fecal immunochemical test. *Clin Gastroenterol Hepatol* 2013;11(7):832-838 e1-2.
11. Ahlquist DA, Sargent DJ, Loprinzi CL, et al. Stool DNA and occult blood testing for screen detection of colorectal neoplasia. *Ann Int Med* 2008;149(7):441-450, W81.
12. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. *N Engl J Med* 2004;351(26):2704-2714.
13. Lieberman DA, Weiss DG. One-time screening for colorectal cancer with combined fecal occult-blood testing and examination of the distal colon. *N Engl J Med* 2001;345(8):555-560.
14. Lee JK, Liles EG, Bent S, et al. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. *Ann Int Med* 2014;160(3):171.
15. Bosch LJ, Oort FA, Neerincx M, et al. DNA methylation of phosphatase and actin regulator 3 detects colorectal cancer in stool and complements FIT. *Cancer Prev Res* 2012;5(3):464-72.
16. Berger DH. Plasmin/plasminogen system in colorectal cancer. *World J Surg* 2002;26(7):767-771.
17. Mueller BM, Yu YB, Laug WE. Overexpression of plasminogen activator inhibitor 2 in human melanoma cells inhibits spontaneous metastasis in scid/scid mice. *Proc Natl Acad Sci U S A* 1995;92(1):205-209.
18. Su CY, Liu YP, Yang CJ, et al. Plasminogen Activator Inhibitor-2 Plays a Leading Prognostic Role among Protease Families in Non-Small Cell Lung Cancer. *PloS One* 2015;10(7):e0133411.
19. Yoshino H, Endo Y, Watanabe Y, et al. Significance of plasminogen activator inhibitor 2 as a prognostic marker in primary lung cancer: association of decreased plasminogen activator inhibitor 2 with lymph node metastasis. *Br J Cancer* 1998;78(6):833-839.
20. Duggan C, Kennedy S, Kramer MD, et al. Plasminogen activator inhibitor type 2 in breast cancer. *Br J Cancer* 1997;76(5):622-627.
21. Chambers SK, Ivins CM, Carcangiu ML. Expression of plasminogen activator inhibitor-2 in epithelial ovarian cancer: a favorable prognostic factor related to the actions of CSF-1. *Int J Cancer* 1997;74(6): 571-575.
22. Nordengren J, Fredstorp Lidebring M, Bendahl PO, et al. High tumor tissue concentration of plasminogen activator inhibitor 2 (PAI-2) is an independent marker for shorter progression-free survival in patients with early stage endometrial cancer. *Int J Cancer* 2002;97(3):379-385.
23. Ganesh S, Sier CF, Griffioen G, et al. Prognostic relevance of plasminogen activators and their inhibitors in colorectal cancer. *Cancer Res* 1994;54(15):4065-4071.
24. Valiente M, Obenauf AC, Jin X, et al. Serpins promote cancer cell survival and vascular co-option in brain metastasis. *Cell* 2014;156(5):1002-1016.

25. Ahlquist DA, Zou H, Domanico M, et al. Next-generation stool DNA test accurately detects colorectal cancer and large adenomas. *Gastroenterology* 2012;142(2):248-256; quiz e25-e26.
26. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med* 2014;370(14):1287-1297.
27. Lieberman DA, Rex DK, Winawer SJ, et al. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology* 2012;143(3):844-857.
28. Sillars-Hardebol AH, Carvalho B, van Engeland M, et al. The adenoma hunt in colorectal cancer screening: defining the target. *J Pathol* 2012;226(1):1-6.
29. Winawer SJ, Zauber AG, Fletcher RH, et al. Guidelines for colonoscopy surveillance after polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer and the American Cancer Society. *Gastroenterology* 2006;130(6):1872-1885.
30. A stool DNA test (Cologuard) for colorectal cancer screening. *Med Lett Drugs Ther* 2014;56(1453): 100-101.
31. Ayala GE, Dai H, Powell M, et al. Cancer-related axonogenesis and neurogenesis in prostate cancer. *Clin Cancer Res* 2008;14(23):7593-7603.
32. Albo D, Akay CL, Marshall CL, et al. Neurogenesis in colorectal cancer is a marker of aggressive tumor behavior and poor outcomes. *Cancer* 2011;117(21):4834-4845.
33. Ondicova K, Mravec B. Role of nervous system in cancer aetiopathogenesis. *Lancet Oncol* 2010;11(6):596-601.
34. Church TR, Wandell M, Lofton-Day C, et al. Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. *Gut* 2014;63(2):317-325.
35. Lee BB, Lee EJ, Jung EH, et al. Aberrant methylation of APC, MGMT, RASSF2A, and Wif-1 genes in plasma as a biomarker for early detection of colorectal cancer. *Clin Cancer Res* 2009;15(19):6185-6191.
36. Lange CP, Campan M, Hinoue T, et al. Genome-scale discovery of DNA-methylation biomarkers for blood-based detection of colorectal cancer. *PloS One* 2012;7(11):e50266.
37. Roperch JP, Incitti R, Forbin S, et al. Aberrant methylation of NPY, PENK, and WIF1 as a promising marker for blood-based diagnosis of colorectal cancer. *BMC Canc*er 2013;13:566.
38. Pedersen SK, Baker RT, McEvoy A, et al. A two-gene blood test for methylated DNA sensitive for colorectal cancer. *PloS One* 2015;10(4):e0125041.
39. Ahlquist DA, Taylor WR, Mahoney DW, et al. The stool DNA test is more accurate than the plasma septin 9 test in detecting colorectal neoplasia. *Clin Gastroenterol Hepatol* 2012;10(3):272-277 e1.
40. Bailey VJ, Zhang Y, Keeley BP, et al. Single-tube analysis of DNA methylation with silica superparamagnetic beads. *Clin Chem* 2010;56(6):1022-1025.
41. Bailey VJ, Easwaran H, Zhang Y, et al. MS-qFRET: a quantum dot-based method for analysis of DNA methylation. *Genome Res* 2009;19(8):1455-1461.
42. Guzzetta AA, Pisanic Ii TR, Sharma P, et al. The promise of methylation on beads for cancer detection and treatment. *Exp Rev Mol Diagn* 2014;14(7):845-852.
43. Kern SE. Why your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures. Cancer Res 2012;72(23):6097-6101.
44. Heyn H, Esteller M. DNA methylation profiling in the clinic: applications and challenges. Nature reviews. Genetics 2012;13(10):679-692.
45. Young GP, Senore C, Mandel JS, et al. Recommendations for a step-wise comparative approach to the evaluation of new screening tests for colorectal cancer. Cancer 2016;122(6):826-839.

Summary

Colorectal cancer (CRC) is one of the most preventable cancers in the world, with over 1,2 million newly diagnosed patients annually. Secondary prevention of CRC, by screening for early disease and precursor lesions, reduces incidence, mortality, morbidity and treatment costs. Over the past two decades, many countries have implemented screening for CRC. Imaging techniques such as flexible sigmoidoscopy, colonoscopy, and CT-colonography and non-invasive stool tests such as guaiac fecal occult blood test (gFOBT) and fecal immunochemical test (FIT) are being used as screening modalities for CRC. Although the sensitivity and specificity of imaging techniques is high, drawbacks are the high costs, complication risks, patient discomfort and requirement of experienced endoscopists. While gFOBT/FIT tests are noninvasive, cheap and easy to use, sensitivity rates for detecting CRCs and advanced lesions are inferior compared to colonoscopy. Additionally, these tests have been shown to obtain high false positive rates, together underscoring the need of sensitive and specific molecular marker tests for early detection of CRC as described in **chapter 1**.

The **aim** of this thesis was to evaluate novel, sensitive and specific promoter methylation markers for noninvasive early detection of CRC and explore the biological function of the identified biomarkers in CRC carcinogenesis.

In **chapter 2**, we analyzed 102 CRCs and 230 noncancerous controls for *GATA4* and *GATA5* promoter methylation which was detected in 70% and 79% of the CRC tissues, respectively. The frequency of *GATA4* and *GATA5* methylation in normal colon mucosa tissue was 6 and 13% respectively. Adenoma tissues collected from CRC patients and noncancerous individuals exhibited methylation frequencies of 33-50% and 41-62%, for *GATA4* and *GATA5* respectively, suggesting that *GATA4/5* methylation is an early event in CRC carcinogenesis. To explore the biomarker potential of the most specific GATA marker, *GATA4* was further analyzed in two series of stool samples collected from two independent series of CRC patients (n=28 and n=47) and colonoscopy negative controls (n=28 and n=30). Using quantitative MSP, a sensitivity of 51-71% and specificity of 84-93% for CRC detection was obtained. In addition, we demonstrated that GATA4 and GATA5 proteins exhibit tumor suppressor characteristics. Transfection of *GATA4* and *GATA5* in CRC cell lines significantly suppressed proliferation, migration, invasion, and anchorage-independent growth of CRC cells. A large amount of literature describing the role of GATA transcription factors is available. In **chapter 3**, these studies are summarized in a comprehensive review discussing the function of GATA transcription factors in development and disease.

Another candidate gene which we identified using epigenome-wide screening approaches was the *N-myc downstream regulated gene 4* (*NDRG4).* We examined promoter methylation of *NDRG4* in CRC and adenoma tissues, as described in **chapter 4**. Promoter methylation was detected in two independent series of CRC tissue

with a sensitivity of 86% (n=83) and 70% (n=184) and a specificity of 96%. As described for *GATA4* and *GATA5* methylation, *NDRG4* methylation was also detected in 55% of adenomas from 62 CRC patients and 66% of adenomas from 22 noncancerous controls. Quantitative MSP was used to study promoter methylation of *NDRG4* in stool DNA of CRC patients and healthy, colonoscopy negative controls of 50 years or older in two independent series. The first series (CRCs: n=28, controls: n=45) showed a sensitivity of 71% and specificity of 84% for CRC detection and validation to confirm the test performance in a second series (CRCs: n=47, controls: n=30) yielded a sensitivity of 51% and specificity of 93%. Overexpression of NDRG4 in human CRC cell lines revealed several tumor suppressive features compared to control transfectants, i.e. reduced colony formation, proliferation and invasion.

Since data about expression and function of *NDRG4* is limited, we explored the expression profile in human and mice by immunohistochemistry, *in situ* mRNA hybridization and Western blotting in **chapter 5**. The *NDRG4* gene is a member of the N-myc downregulated gene family and expression was reported in the central nervous system and heart, which was confirmed in this study. However, in the heart, NDRG4 expression was not observed in the cardiomyocytes, but was restricted to specialized subendocardial Purkinje fibers. Furthermore, NDRG4 protein expression was shown in the peripheral nervous system throughout the murine body. Surprisingly in the gastrointestinal tract, no NDRG4 expression was shown in the epithelial cells, but NDRG4 expression was confined to the enteric nervous system (ENS), including the submucosal (Meissner's) plexus and myenteric (Auerbach's) plexus. *In situ* hybridization confirmed NDRG4 expression in the ENS and immunofluorescence assays showed restriction of NDRG4 to neurons, as NDRG4 co-localized with the pan-neuronal marker HuC/D but never co-localized with the glia marker GFAP.

To explore the performance of candidate methylation markers in blood, *NDRG4*, *GATA5*, *SYNE1* and *FOXE1* methylation was analyzed in plasma of 154 CRC patients and 444 endoscopy negative controls in **chapter 6**. *GATA5* and *NDRG4* methylation was detected in 18% and 27% of CRC patients, respectively. *SYNE1* and *FOXE1* methylation frequencies of 47% and 46% were observed with specificities of 96% and 93%, respectively. When *SYNE1* and *FOXE1* are combined, the sensitivity increased to 58% with a specificity of 91%. Analysis of *SYNE1* and *FOXE1* in a panel showed a clear stage dependency with sensitivities ranging from 37% in stage I to 100% in stage IV CRC. Functional assays have been performed in order to investigate the biological function of SYNE1 and FOXE1 proteins using transfected CRC cell lines. No significant effect on migration or invasion was observed upon *SYNE1* transfection in CRC cell lines. Surprisingly, *SYNE1* transfectants showed an increased proliferation compared to control transfectants. Potential suppressive features of FOXE1 were shown, as upregulation in CRC cell lines inhibits colony formation.

Finally, in **chapter 7**, the significance of the results obtained in this thesis is discussed. Our findings, as well as the available literature, suggest a better performance of stool-based DNA tests compared to blood-based tests. Partly, this can be explained by the dependency of blood-borne spread of tumor cells predominantly occurring in a late CRC stage in which patients already have a worse prognosis. The data in this thesis also indicate that *NDRG4* performs best of the analyzed stool-based epigenetic markers in CRC detection, which is confirmed by others. Finally, further developments to improve the clinical and analytical sensitivity, such as the identification of novel and/or complementary markers and implementation of cutting-edge techniques in order to increase the sensitivity for CRC detection, are discussed.

Valorization

Colorectal cancer (CRC) is a major burden on the health care system with over 1,4 million newly diagnosed patients and almost 700,000 deaths annually.[1] Because of these numbers, the global economic burden is substantial, with an estimated $US14-22 billion each year. Most of the costs occur in advanced stage CRC and are related to hospitalization, chemo- and radiotherapy, treatment of related side-effects and supportive care.[2] In the Netherlands, CRC is one of the most frequently diagnosed malignancies with over 15,000 new cases and over 5,000 cancer-related deaths per year.[3] In 2011, €488 million were spent on health care of CRC patients in the Netherlands.[4,5] In order to decrease CRC rates and thus also the economic as well as the social burden, it is generally accepted that the most effective strategy to manage the disease is early detection.[6]

In addition to invasive screening methods, i.e. colonoscopy and sigmoidoscopy, several non-invasive methods have been developed such as the fecal occult blood test (FOBT) and fecal immunochemical hemoglobin test (FIT). It has already been studied that population-based screening using FOBT leads to a reduction in CRC mortality rates.[7-10] Since FIT achieves higher participation and CRC detection rates[11-13], population-based screening with this test is expected to improve the mortality rates even more. While FIT is cost-effective and is expected to reduce CRC incidence and mortality, its performance can be improved since the test still misses approximately 20% of CRCs and up to 70% of advanced neoplasia.[11,14,15] Lowering the cut-off value increases the detection of advanced neoplasia, but results in a decrease of the positive predictive value leading to unnecessary patient anxiety and costs of follow up examinations.

In order to improve CRC screening , we identified several promoter methylation biomarkers in blood and/or feces. In Chapter 2, the potential of promoter methylation detection of *GATA4* in fecal DNA is investigated, reaching sensitivities and specificities of 51-71% and 84-93%, respectively. Additionally, detection of *GATA5* promoter methylation using blood-based assays as described in chapter 6, yielded a methylation frequency of only 18% with a specificity 99%. *SYNE1* and *FOXE1* obtained respectively detection rates of 47% and 46% with specificities of 96% and 93%. Combining *SYNE1* and *FOXE1* increased the sensitivity to 58% with a minimal decrease of the specificity to 91%. Though these results are promising, the reached sensitivities and specificities are not sufficient for incorporation into a clinical setting. Due to improved fecal and blood DNA isolation and detection techniques, these results might improve in the future. In chapter 4, we showed that detection of *NDRG4* promoter methylation in stool samples identified 53-61% of CRC patients, whereas almost no false positive results were obtained. Before incorporation of a novel biomarker into a screening program can be considered, an optimal sensitivity and specificity should be reached. In addition, extensive validation within the intended target population to confirm the initial results is required.[16] Combining sensitivity and specificity, *NDRG4* is one of the best single early

detection methylation markers published so far. This finding has been validated in independent studies, demonstrating its clinical potential.[17-20] *NDRG4* methylation as a diagnostic marker for CRC has been patented by our group and biomarker company MDxHealth (Irvine, USA, www.mdxhealth.com) and was licensed to Exact Sciences (Madison, USA, www.exactsciences.com), a molecular diagnostics company developing a molecular maker test for CRC. Exact Sciences incorporated *NDRG4* in their multi-marker molecular diagnostic CRC screening test called Cologuard®, which includes detection of *KRAS* mutations and *NDRG4* and *BMP3* promoter methylation together with a human hemoglobin immunoassay.[16] The United States Food and Drug administration (FDA) has approved Cologuard® to screen an average-risk adult population of ≥50 years old for CRC.[21] Additionally, many health insurance companies in the United States of America decided to cover the costs of Cologuard®.

As in many other countries, a population-based CRC screening program has been introduced in the Netherlands since 2014, in which individuals of 55-75 years or older are offered an immunochemical fecal occult blood test (FIT). If positive, a follow-up colonoscopy will be performed. The costs are funded by the national budget and therefore paid with tax revenues. This does not include the costs of the colonoscopy in the case of a positive FIT, which is paid by the health insurance of the identified individual. Comparison of Cologuard® with FIT in a large prospective study, showed a significantly higher detection rate of 85% and a specificity of 95% with the Cologuard® test, while the FIT achieved less false positive results.[19] Although the sensitivity of the Cologuard® is promising, several adaptations are required before the test can be incorporated in countries such as the Netherlands, in which the logistic system is designed for screenees receiving a FIT which only requires a minimal amount of stool and is feasible for at-home testing. Using Cologuard®, whole stool samples have to be transported to a laboratory which is not only a logistic and economic burden but also requires adequate transportation material and preservation buffers. Improvement of Cologuard® can be obtained by further development of at-home testing on small amounts of stool. Several factors influence the amount of DNA and its quality, such as sample collection, storage buffers and DNA isolation. In order to increase the yield of high quality DNA, these factors should be optimized.

In conclusion, one of the biomarkers identified in this thesis has been incorporated into a commercial biomarker-assay that detects more CRCs when compared to FIT. If this biomarker assay will be introduced in more CRC screening programs, we expect a decrease in CRC morbidity and mortality and consequently a decrease of the social and economic burden.

References

1. Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. *CA Cancer J Clin* 2015. 65(2):87-108.
2. Frazier AL, Colditz GA, Fuchs CS, et al., *C*ost-effectiveness of screening for colorectal cancer in the general population. *JAMA* 2000;284(15):1954-1961.
3. Integraal Kankercentrum Nederland. 2015; Available from: www.iknl.nl.
4. Database RKvZ. 2013; Available from: www.kostenvanziekten.nl.
5. Jansman FG, Postma MJ, and Brouwers JR. Cost considerations in the treatment of colorectal cancer. *PharmacoEconomics* 2007;25(7):537-562.
6. Edwards BK, Ward E, Kohler BA, et al. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer* 2010;116(3):544-573.
7. Hardcastle, J.D., Chamberlain, J.O., Robinson, M.H., et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet* 1996;348(9040):1472-1477.
8. Mandel JS, Bond JH, Church TR, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. *N Engl J Med* 1993;328(19):1365-1371.
9. Kronborg O, Jorgensen OD, Fenger C, et al. *Randomized study of biennial screening with a faecal occult blood test: results after nine screening rounds. Scand J Gastroenterol* 2004;39(9):846-851.
10. Hewitson P, Glasziou P, Watson E, et al. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. *Am J Gastroenterol* 2008;103(6):1541-1549.
11. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. *Gut* 2015;64(10):1637-1649.
12. Hol L, van Leerdam ME, van Ballegooijen M, et al. Screening for colorectal cancer: randomised trial comparing guaiac-based and immunochemical faecal occult blood testing and flexible sigmoidoscopy. Gut 2010;59(1):62-68.
13. van Rossum LG, van Rijn AF, Laheij RJ, et al. Random comparison of guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. *Gastroenterology* 2008;135(1):82-90.
14. Lansdorp-Vogelaar I, Knudsen AB, and Brenner H. Cost-effectiveness of colorectal cancer screening. *Epidemiol Rev* 2011;33:88-100.
15. Chen LS, Liao CS, Chang SH, et al. Cost-effectiveness analysis for determining optimal cut-off of immunochemical faecal occult blood test for population-based colorectal cancer screening (KCIS 16). *J Med Screen* 2007;14(4):191-199.
16. Ahlquist DA, Zou H, Domanico M, et al. Next-generation stool DNA test accurately detects colorectal cancer and large adenomas. *Gastroenterology* 2012;142(2):248-56; quiz e25-6.
17. Ahlquist DA, Zou H, Domanico M, et al. Next-generation stool DNA test accurately detects colorectal cancer and large adenomas. *Gastroenterology* 2012;142(2):248-56.
18. Ahlquist DA, Taylor WR, Mahoney DW, et al. The stool DNA test is more accurate than the plasma septin 9 test in detecting colorectal neoplasia. *Clin Gastroenterol Hepatol* 2012;10(3):272-277 e1.
19. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med* 2014;370(14):1287-1297.
20. Lidgard GP, Domanico MJ, Bruinsma JJ, et al. Clinical performance of an automated stool DNA assay for detection of colorectal neoplasia. *Clin Gastroenterol Hepatol* 2013;11(10):1313-1318.
21. A stool DNA test (Cologuard) for colorectal cancer screening. *JAMA* 2014;312(23):2566.

Dankwoord

Zo, we zijn aangekomen bij het meest gelezen hoofdstuk van een proefschrift, maar verder voel ik geen druk hoor…

Als je maar liefst 13 jaar over je proefschrift doet is het een illusie om te denken dat je iedereen in je dankwoord kwijt kan zonder iemand te vergeten. Daarnaast zal het dankwoord dan waarschijnlijk de helft van het proefschrift innemen want jullie zijn met zovelen, letterlijk 'teveel om op te noemen'. Zo zijn er mensen die in directe zin zeer betrokken zijn geweest, zoals de begeleiding rondom het onderzoek, het kritisch bekijken van manuscripten en de ondersteuning op het lab. Opbeurende en moedgevende woorden, schoppen onder m'n kont, complimenten en steun heb ik regelmatig mogen ontvangen. Daarnaast zijn er velen die in de directe werkomgeving voor een uitstekende ('misschien niet altijd leidend tot'-werk….)sfeer zorgen. Vele hilarische momenten op het research maar ook klinisch lab, in de AIO en AIOS kamers, uitsnijkamer, obductieruimte (ja, ook daar!) en op het secretariaat staan mij bij. Indirect zijn de leuke werk gerelateerde contacten, uitjes en bezoeken van cursussen, congressen maar ook de memorabele feestjes zoals verjaardagen, bruiloften, en natuurlijk promoties, van niet te onderschatte waarde!

Naast het werk is de bijdrage van de privé-omgeving minstens zo belangrijk bij het tot stand komen van een proefschrift. Zo zijn er een heel aantal zeer dierbare personen in mijn leven die al heel lang mijn doen en laten volgen en mij met bergen geduld steunen in al mijn gestelde doelen. Familieweekenden, uitstapjes, gezellige etentjes en vakanties met familie en vrienden maken het leven niet alleen veel leuker maar zorgen ook voor het benodigde relativeringsvermogen. Er zijn veel lieve mensen in onze omgeving die bereid waren (en zijn, hoop ik ☺) om op "De Monsters" te passen zodat Hans en ik samen naar bijvoorbeeld een ouderavond of gebarentaalcursus konden, of gewoon eens lekker met z'n tweeën er tussenuit (sorry jongens…). Of die in een onbewaakt ogenblik ervoor zorgden dat de strijk, of dergelijk terugkerend leed, ineens was gedaan. Dit zorgde voor de benodigde ruimte en rust in mijn hoofd om op andere momenten ruimte voor het werken aan het onderzoek te hebben. Tenslotte ook de mensen die in de wandelgangen altijd maar weer met oprechte belangstelling vroegen hoe het (onder andere) met het onderzoek ging en daarbij lieten blijken dat ze mij geloofden dat er ooit (niet te geloven maar het is echt zover!!!) een feestje zou komen.

Kortom, er zijn heel veel fijne, lieve en belangrijke mensen die in de loop van de tijd hebben bijgedragen aan het totstandkoming van dit boekje op wat voor een manier dan ook, die ik bij deze uit de grond van mijn hart wil bedanken! Zonder jullie had dit proefschrift er nooit gelegen!

Curriculum vitae

Curriculum vitae

Marjolein Hendrina Francisca Maria Lentjes werd geboren op 26 januari 1977 te Nijmegen. Na de lagere school ging zij met havo-vwo niveau op zak, tegen alle goed bedoelde adviezen in, naar de Lagere land- en tuinbouwschool in Nijmegen. Tijdens deze opleiding verschoof haar belangstelling voor bloemsierkunst naar plantenveredeling waarvoor een laboratoriumopleiding vereist was. Na het behalen van haar diploma werd zij, via een verkorte havo-opleiding, in 1996 aangenomen op de Hogere Laboratorium Opleiding aan de Hogeschool van Arnhem en Nijmegen te Nijmegen. In de loop van deze opleiding werd de interesse voor de medische wereld gewekt en studeerde zij in 2000 af als cytologisch analist. Echter was er tijdens de afsluitende stage op de afdeling Pathologie in het Maastricht Universitair Medisch Centrum (MUMC) een onbedwingbare ambitie ontstaan om klinisch patholoog te worden. Na één jaar deels gewerkt te hebben als IVF-analist en deels als research analist bij de afdeling Obstetrie & Gynaecologie aan het MUMC werd zij ingeloot voor de studie Geneeskunde aan de Universiteit van Maastricht. Haar co-assistentschappen werden onder andere volbracht op de afdeling KNO van het Hospital Universitário Walter Cantídio te Fortaleza in Brazilië. Gedurende de studie Geneeskunde heeft zij in eerste instantie op de afdeling Pathologie van het MUMC als research analist moleculair biologische technieken bij Prof. dr. A. de Goeij gewerkt. In 2003 startte zij als AIO onder supervisie van Prof. dr. A. de Bruïne en Prof. dr. M. van Engeland op dezelfde afdeling Pathologie. Na het behalen van het artsexamen in december 2008 werd zij in januari 2009 aangenomen als AIOS Pathologie aan het MUMC waarbij de perifere stages in het Atrium Medisch Centrum in Heerlen en het VieCuri ziekenhuis in Venlo werden voltooid. Tijdens deze opleiding werd het onderzoeksproject als AIO vervolgd wat uiteindelijk heeft geleid tot dit proefschrift. In april 2015 werd de opleiding Pathologie afgerond en sinds 1 mei 2015 is ze als klinisch patholoog werkzaam bij de stichting PAMM in Eindhoven.

List of publications

List of publications

Lentjes MH, Vaes N, Gijbels MJ, Rademakers G, Daenen KL, Wouters KA, Geuzens A, Qu X, Steinbusch HP, Rutten B, Baldwin S, Sharkey KA, Hofstra RM, van Engeland M, Vanden Berghe P, Melotte V. NDRG4 is predominantly expressed in neurons of the central and peripheral nervous system and restricted to enteric neurons of the mouse and human intestinal tract. *Submitted*

Lentjes MH, Niessen HE, Akiyama Y, de Bruïne AP, Melotte V, van Engeland M. The emerging role of GATA transcription factors in development and disease. Expert Rev Mol Med. 2016 Mar 8;18:e3.

van Osch FH, Voets AM, Schouten LJ, Gottschalk RW, Simons CC, van Engeland M, **Lentjes MH**, van den Brandt PA, Smeets HJ, Weijenberg MP. Mitochondrial DNA copy number in colorectal cancer: between tissue comparisons, clinicopathological characteristics and survival. Carcinogenesis. 2015 Dec;36(12):1502-10.

Melotte V*, Yi JM*, **Lentjes MH**, Smits KM, Van Neste L, Niessen HE, Wouters KA, Louwagie J, Schuebel KE, Herman JG, Baylin SB, van Criekinge W, Meijer GA, Ahuja N*, van Engeland M*. Spectrin repeat containing nuclear envelope 1 and forkhead box protein E1 are promising markers for the detection of colorectal cancer in blood. Cancer Prev Res (Phila). 2015 Feb;8(2):157-64.

Wilop S, van Gemmeren TB, **Lentjes MH**, van Engeland M, Herman JG, Brümmendorf TH, Jost E, Galm O. Methylation-associated dysregulation of the suppressor of cytokine signaling-3 gene in multiple myeloma. Epigenetics. 2011 Aug;6(8): 1047-52.

Melotte V, **Lentjes MH**, van den Bosch SM, Hellebrekers DM, de Hoon JP, Wouters KA, Daenen KL, Partouns-Hendriks IE, Stessels F, Louwagie J, Smits KM, Weijenberg MP, Sanduleanu S, Khalid-de Bakker CA, Oort FA, Meijer GA, Jonkers DM, Herman JG, de Bruïne AP, van Engeland M. N-Myc downstream regulated gene 4 (NDRG4): a candidate tumor suppressor gene and potential biomarker for colorectal cancer. J Natl Cancer Inst. 2009 Jul 1;101(13):916-27.

Hellebrekers DM*, **Lentjes MH***, van den Bosch SM, Melotte V, Wouters KA, Daenen KL, Smits KM, Akiyama Y, Yuasa Y, Sanduleanu S, Khalid-de Bakker CA, Jonkers D, Weijenberg MP, Louwagie J, van Criekinge W, Carvalho B, Meijer GA, Baylin SB, Herman JG, de Bruïne AP, van Engeland M. GATA4 and GATA5 are potential tumor suppressors and biomarkers in colorectal cancer. Clin Cancer Res. 2009 Jun 15;15(12):3990-7.

de Vogel S, van Engeland M, Lüchtenborg M, de Bruïne AP, Roemen GM, **Lentjes MH**, Goldbohm RA, van den Brandt PA, de Goeij AF, Weijenberg MP. Dietary folate and APC mutations in sporadic colorectal cancer. J Nutr. 2006 Dec;136(12):3015-21.

Derks S, Postma C, **Lentjes MH**, van Engeland M, Meijer G.A. Het ontstaan van dikke darm kanker. Tijdschrift Kanker 2. 2005: 34-7

Lüchtenborg M, Weijenberg MP, de Goeij AF, Wark PA, Brink M, Roemen GM, **Lentjes MH**, de Bruïne AP, Goldbohm RA, van 't Veer P, van den Brandt PA. Meat and fish consumption, APC gene mutations and hMLH1 expression in colon and rectal cancer: a prospective cohort study (The Netherlands). *Cancer Causes Control.* 2005 Nov;16(9):1041-54.

Brink M, Weijenberg MP, de Goeij AF, Roemen GM, **Lentjes MH**, de Bruïne AP, Goldbohm RA, van den Brandt PA. Meat consumption and K-ras mutations in sporadic colon and rectal cancer in The Netherlands Cohort Study. *Br J Cancer*. 2005 Apr 11;92(7):1310-20.

Brink M, Weijenberg MP, de Goeij AF, Roemen GM, **Lentjes MH**, de Bruïne AP, van Engeland M, Goldbohm RA, van den Brandt PA. Dietary folate intake and K-ras mutations in sporadic colon and rectal cancer in The Netherlands Cohort Study. Int J Cancer. 2005 May 1;114(5):824-30.

Derks S*, **Lentjes MH***, Hellebrekers DM, de Bruïne AP, Herman JG, van Engeland M. Methylation-specific PCR unraveled. Cell Oncol. 2004;26(5-6):291-9.

Brink M, Weijenberg MP, de Goeij AF, Schouten LJ, Koedijk FD, Roemen GM, **Lentjes MH**, de Bruïne AP, Goldbohm RA, van den Brandt PA. Fat and K-ras mutations in sporadic colorectal cancer in The Netherlands Cohort Study. *Carcinogenesis*. 2004 Sept;25(9): 1619-28.

Lüchtenborg M, Weijenberg MP, Roemen GM, de Bruïne AP, van den Brandt PA, **Lentjes MH**, Brink M, van Engeland M, Goldbohm RA, de Goeij AF. APC mutations in sporadic colorectal carcinomas from The Netherlands cohort study. *Carcinogenesis*. 2004 Jul;25(7): 1219-26.

Brink M, de Goeij AF, Weijenberg MP, Roemen GM, **Lentjes MH**, Pachen MM, Smits KM, de Bruïne AP, Goldbohm RA, van den Brandt PA. K-ras oncogene mutations in sporadic colorectal cancer in The Netherlands Cohort Study. *Carcinogenesis.* 2003 Apr;24(4): 703-710.

(*equal contribution)